

Structural Inference in Epochal Evolution

PHY256B Class Project

Anastasia Belenkii^{1*} Mikhail Semaan² James P. Crutchfield²

¹University of California at Davis, One Shields Avenue, Davis, CA 95616

²Complexity Sciences Center and Department of Physics and Astronomy, University of California at Davis, One Shields Avenue, Davis, CA 95616

June 12, 2022

A simple evolutionary system is examined. Building off of [6], we investigate the information-theoretic properties of a "Royal Road" Genetic Algorithm that optimizes for the greatest number of length- K blocks of 1s in a binary string. By measuring the statistical complexity and entropy rate of the fittest bit string at each generation, we find that not all epochs are created equal, with some periods of stability during the underlying evolutionary process having markedly different informational properties that may hint at specific trajectories and even causal explanations not visible from observing the population's overall fitness alone.

1 Introduction

1.1 Background

The reader is expected to be familiar with the basics of computational mechanics, presented in compact form in [1] and [5]. In this paper, our main tools will be the quantities C_μ , the statistical complexity, and $h_\mu(L)$, the entropy rate with a history of length L . h_μ (as implemented in [4]) is defined as:

$$\begin{aligned} H_X(k) &= \langle h_{X,i}(k) \rangle_i \\ &= \sum_{x_i^{(k)}, x_{i+1}} p(x_i^{(k)}, x_{i+1}) \log_2 \frac{p(x_i^{(k)}, x_{i+1})}{p(x_i^{(k)})}. \end{aligned}$$

Representing the "entropy density", or the rate at which the entropy increases as the history length increases. We are not taking the infinite limit here, but instead setting the maximum history to the length of the blocks we are trying to match in the GA.

C_μ is defined as:

*ambelenkii@ucdavis.edu

$$\begin{aligned}
C_\mu(R) &\equiv H[R] \\
&= - \sum_{\rho \in R} P(R = \rho) \log_2 P(R = \rho),
\end{aligned}$$

Where R is the class of a process' causal states (this class is minimally complex and maximally predictive of the process, as shown in the references above). It measures the uncertainty in representation of the process that generated the observed data separate from the uncertainty in the data itself, and thus reflects the amount of memory a process uses to generate symbols.

2 Methods

We used the same initial setup as [6]: A population of binary strings of length $L = N \times K$, where K is the length of a block and N is the number of blocks. We calculate fitness according to the number of matching blocks, and select randomly from the population to create the next generation, weighted according to the fitness. We also subject each bit in the string to a probability q of flipping.

The subsequent portion of our analysis consisted of the following steps: 1. calculate and save the best-fit individual from each generation¹, 2. Infer a unifilar Hidden Markov Model for each fittest individual over a prior (uniform) distribution of all 1-4 state unifilar Hidden Markov Models² [2], and calculate the statistical complexity for each individual. We calculated the entropy rate directly from the bitstrings instead of using HMMs to minimize possible noise from our inference method.

3 Results

To begin, the main figure from [6] was reproduced. It also serves as an important reference point to the following graphs, in order to see where in the progression of the population C_μ and h_μ undergo major changes.

Although we used a run of 4500 iterations for this analysis, we capped our datasets for each quantity at 2475. After this generation, h_μ was exclusively zero. In addition, it was simply too time-consuming to compute more iterations than this for C_μ , and visual observation of the data showed that once the population matched 7 blocks, the series of best fit individuals got "stuck", unable to fixate further matched 1s that would enable it to match the final block. Perhaps with more iterations it would have been possible to exit this epoch, but for brevity's sake we have not included it in this very preliminary analysis.

4 Discussion

We have a mechanistic "story" already in mind: mutation generally either is neutral to or harms the fitness of an individual, but if the mutation matches a block it drastically increases the individual's fitness, leading to a 'founder effect' where there are long periods of stasis punctuated by extremely rapid innovation as the new beneficial mutation becomes fixated

¹it was also possible to use a randomly selected average individual, but we found that the best-fit individuals had more structure overall, making it easier to see trends over time.

²in reality these were samples from a distribution over possible initial states and transition probabilities given the most likely inferred structure

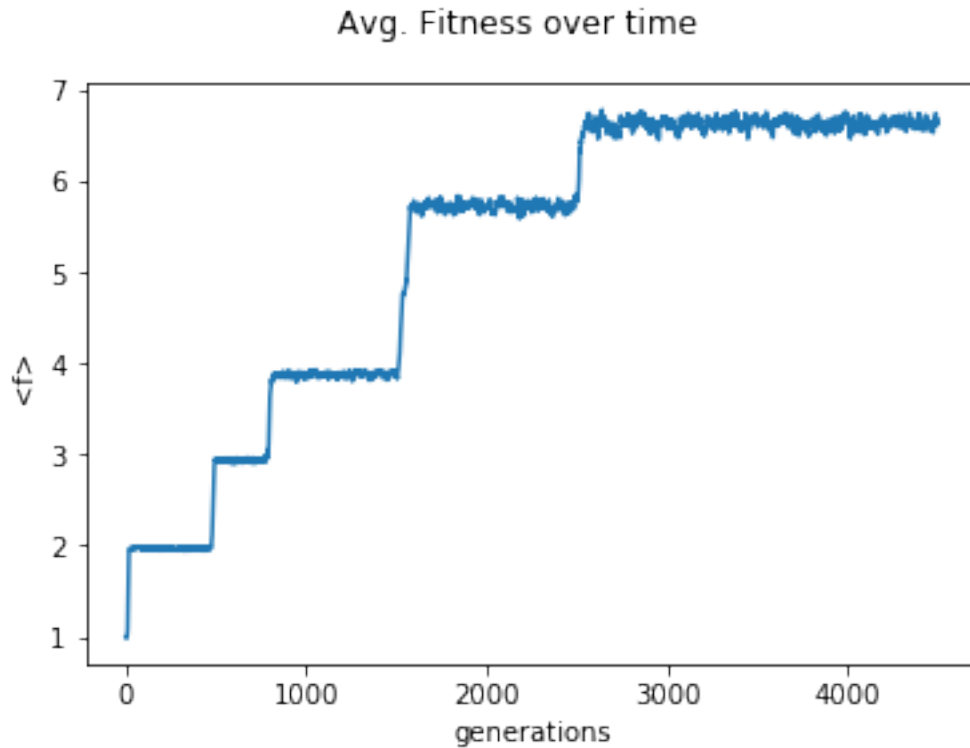


Figure 1: Displaying the Epochal nature of the Royal Road algorithm Main Run, reproducing Fig. 1 of [6]. Params: N and $K = 8$, $M = 500$, $q = 0.01$, for 4500 iterations.

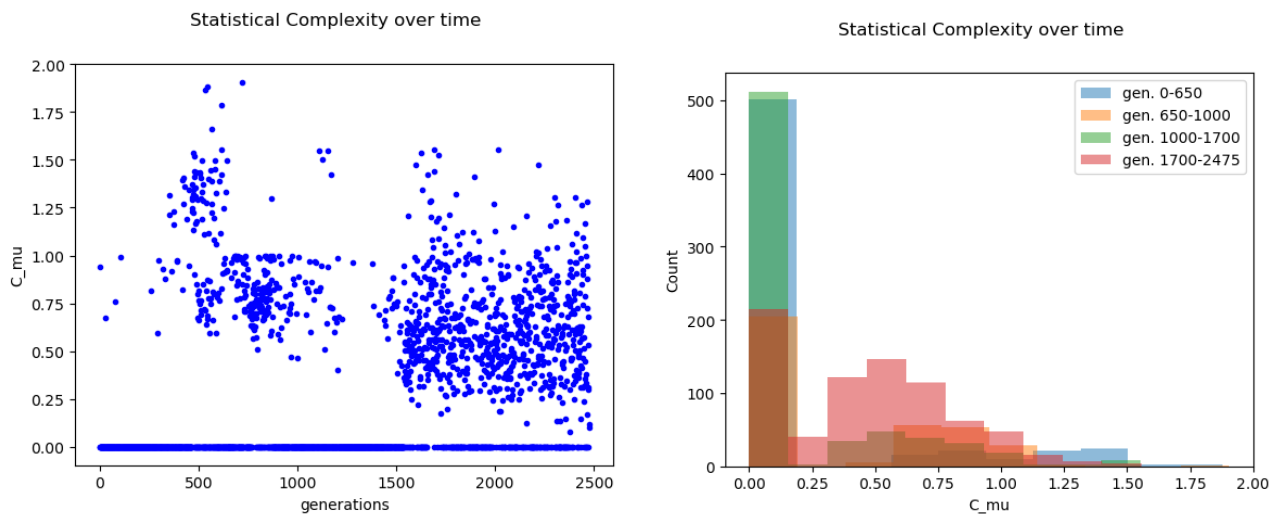


Figure 2: The statistical complexity of randomly sampled machines (i.e., start node and transition probabilities) from the Maximum a. Posteriori Estimate of machine topology, over the course of the run.

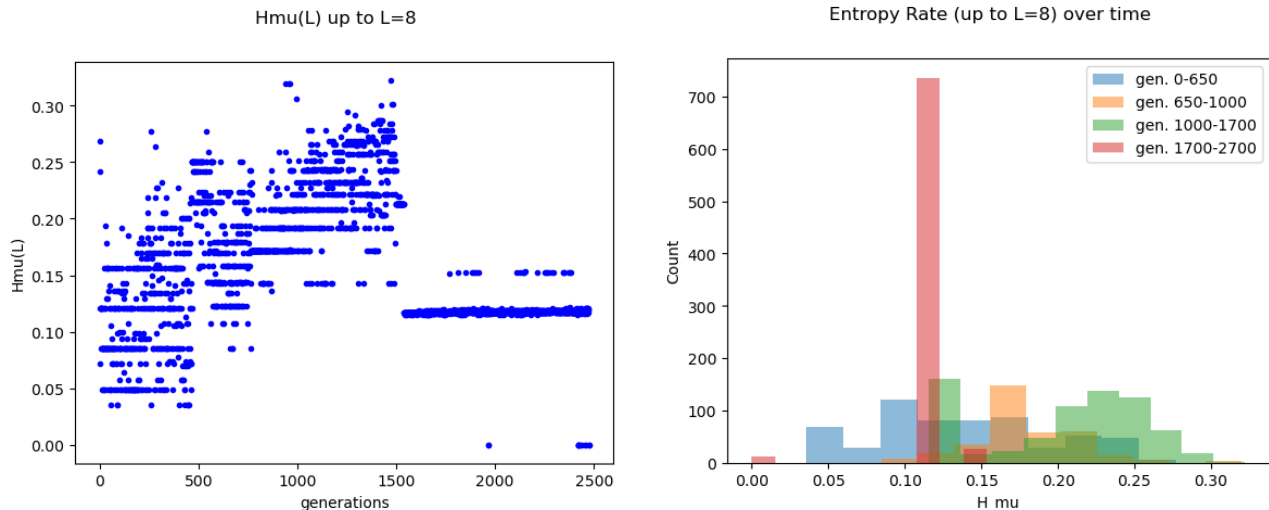


Figure 3: The (finite maximum) entropy rate of a given bitstring at generation g , calculated directly from the bitstring without reference to a given ϵ -machine.

across the population. This is reflected by the findings in [6]: they find that the process has strong metastable points corresponding to the eigenvalues of (a linearized version of) the generation operator $G = M \circ S$, where M and S are the mutation and selection operators, respectively, but that if it starts moving away from these metastable points, it will do so exponentially fast.

We can recognize the epochs in the data we have gathered: the drastic drop in the variance of $H_\mu(L)$ at around generation 1500 in Figure 3 corresponds with the huge increase in average fitness seen in Figure 1– the biggest for the entire run– possibly suggesting a founder effect whereby one sequence completely dominated the entire population. However, there seemed to be more going on:

The variance of Figure 3 clearly increases starting from around the same point in the run, suggesting that the inference process was picking up on some kind of additional structure in that particular sequence. Typically, processes where C_μ would be equal to zero were biased coins with a probability of getting a 1 that more or less corresponded to the population’s average fitness. On visual observation, processes whose most likely topology were not biased coins seemed much more strongly periodic to the naked eye, with fewer stray “ones” or “zeros” breaking up longer (or at least more regular) sequences. The collapse in $h_m u$ to a narrow band of values corresponding to the same period in which there is a larger frequency of $C_\mu = 0$ HMMs is interesting in this light. Also worth noting are the clear bands in the values taken on by h_μ , in both Figure 3 and Figure 4.

Directions for Further Research

Our approach had many limitations, and many aspects of the present study could be extended. Ideally, the same analysis could be fully repeated using randomly sampled average individuals, and compared to the data from the fittest individuals. It would also be a useful sanity check to invest the computational resources to replicate the steps described in the methods section using much longer L values to give the inference process more to work with. Another possible

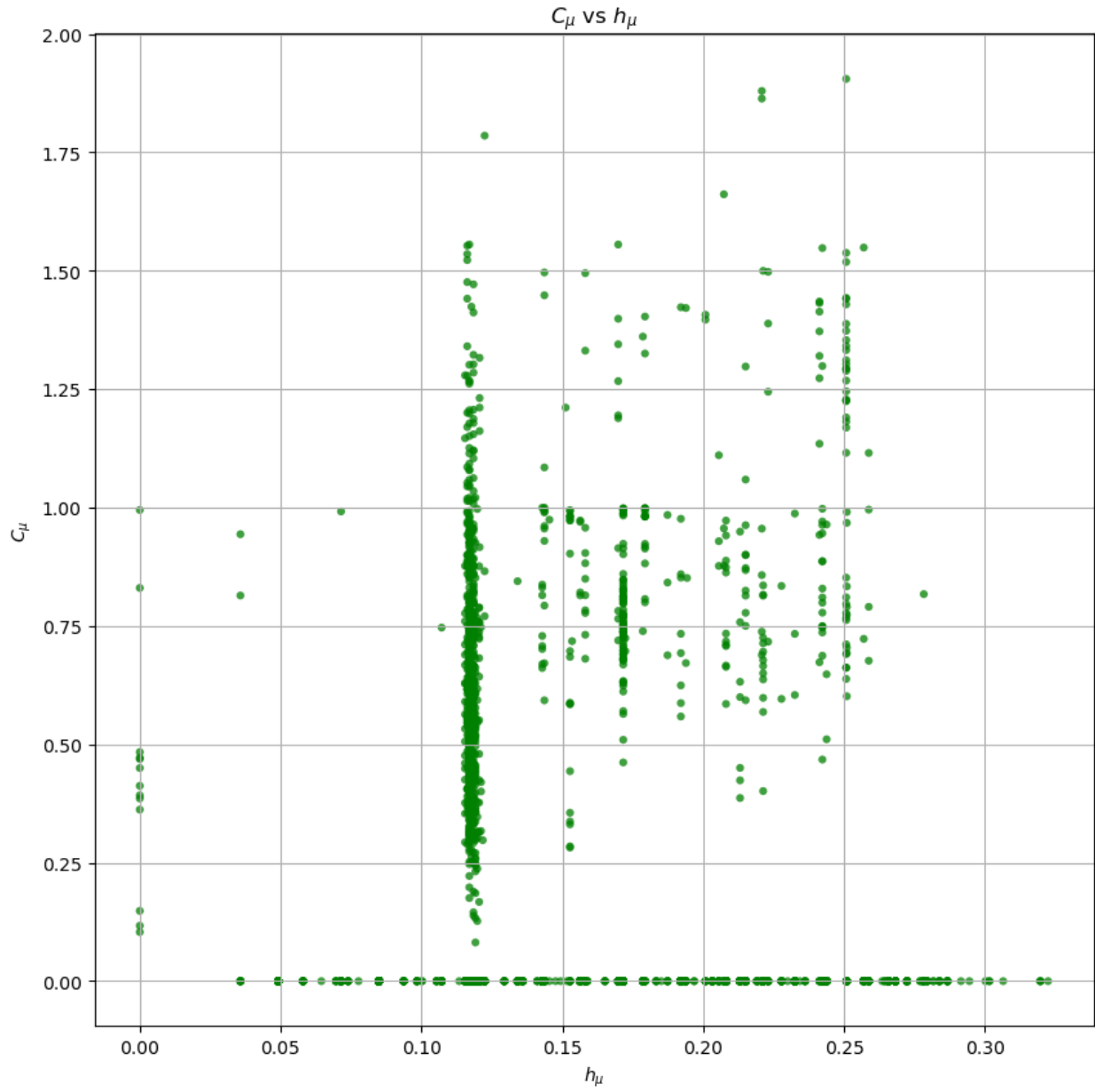


Figure 4: Comparison of C_μ and H_μ across the entire run.

exploration could be focus on the renewal times between fitness spikes, taking inspiration from [3]– and all of that is before even considering other evolutionary process and genetic algorithms.

References

- [1] James P. Crutchfield. The calculi of emergence: computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, 75(1):11–54, August 1994.
- [2] B. D. Johnson, J. P. Crutchfield, C. J. Ellison, and C. S. McTague. Enumerating Finitary Processes. Technical Report arXiv:1011.0036, arXiv, December 2012. arXiv:1011.0036 [nlin, stat] type: article.
- [3] Sarah E. Marzen and James P. Crutchfield. Informational and Causal Architecture of Discrete-Time Renewal Processes. *Entropy*, 17(7):4891–4917, July 2015. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
- [4] Douglas G Moore. Time series measures¶].
- [5] Cosma Rohilla Shalizi and James P. Crutchfield. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics*, 104(3):817–879, August 2001.
- [6] Erik van Nimwegen, James P. Crutchfield, and Melanie Mitchell. Finite Populations Induce Metastability in Evolutionary Search. Working Paper, Santa Fe Institute, August 1996.