“Killing” and “Collapsing”

How varying transition probabilities between states alters the statistical and topological properties of probabilistic epsilon machines

Nikki Sanderson

Department of Mathematics

nfsanderson@ucdavis.edu

Abstract:

In this paper we look at the effects varying the transition probabilities between the states of probabilistic epsilon machines has on the statistical and information-theoretic properties, as well as the topology, of the epsilon machines. Specifically, we measure the changes in entropy rate, statistical complexity, and excess entropy over the transition probability parameter space of the Even Process, Golden Mean Process, Noisy Random Phase Slip Process, and Telescope Process. We discuss the significance of these measurements to the robustness of epsilon machines as models. We then introduce the notions of “killing” and “collapsing”, two topology-altering phenomena that occur as we vary the transition probabilities, and discuss their implications to the study of the space of all probabilistic epsilon machines.

Introduction

Motivation

The motivation behind this research was two-fold. I was curious to see how varying the transition probabilities between states of probabilistic epsilon machines affected statistical measurements of the process being represented; and, more abstractly, I wanted to see if varying transition probabilities would provide hints of structure to the space of all epsilon machines.

Synopsis

We began by parameterizing the transition probabilities between states for the Even Process, Golden Mean Process, Noisy Random Phase Slip Process, and Telescope Process. We then wrote code in the CMPy server to plot the entropy rate, statistical complexity, and excess entropy over the transition probability distribution parameter space. Upon generating the plots, we analyzed the results.

Not only does the data support reason for interest in such plots as a means of measuring model robustness, it illuminates a potential avenue to pursue in future research on the structure of the space of probabilistic epsilon machines.

Striking features found in some plots of entropy rate, statistical complexity, and excess entropy over the transition probability distribution parameter space of these epsilon machines also highlight that topological changes in the epsilon machine representation can occur - via the processes of "killing" and "collapsing" – as a result of varying the transition probability distribution. Where these topological changes occur, what leads to them and how they can be used to study the structure of the space of all epsilon machines are all addressed in this paper.

Background

Model Robustness

Probabilistic epsilon machines were introduced as models of stationary stochastic processes whose states represent equivalence classes of histories that have the same probability distribution over future events. Multiple algorithms have been provided for constructing an epsilon machine representation for a given stochastic process from the data gleaned while measuring the process; in all methods of construction, the transition probabilities between states are derived directly from this data. Thus, variability in the data could lead to variability in the transition probabilities between the states of the epsilon machine.

The entropy rate, **hμ**, of a process is defined below and is interpreted as the intrinsic randomness of a process.

Eq.1:

$$h_\mu(\boldsymbol{\mathcal{S}}) = -\sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \sum_{s\in\mathcal{A},\mathcal{S}'\in\boldsymbol{\mathcal{S}}} T^{(s)}_{\mathcal{S}\mathcal{S}'} \log_2 T^{(s)}_{\mathcal{S}\mathcal{S}'}$$

The statistical complexity, **Cμ**, of a process is defined below and is interpreted as the stored information in a process, or the amount of structure in the process.

Eq.2:

$$C_\mu(\boldsymbol{\mathcal{S}}) = -\sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \log_2 \Pr(\mathcal{S})$$

The excess entropy, **E**, of a process is [ask Ryan how CMPy calculates E] and is interpreted as the amount of information transmitted from the past to the future.

As the entropy rate, statistical complexity, and excess entropy are statistical and information-theoretic properties each defined as a function of the transition probabilities between states, one expects their values to change as the transition probability distribution is varied.

Knowledge of how statistical and information-theoretic properties such as entropy rate, statistical complexity, and excess entropy are affected by changing the transition probabilities of a probabilistic epsilon machine allows one using an epsilon machine representation of a process to provide a bound for these values based on their confidence in their data.

This interpretation of model robustness concerning the sensitivity of statistical measurements to variation of transition probabilities is just one, though one that has the potential to be rigorously defined. The notion of robustness in relation to epsilon machine representations should be explored further in future work, particularly in conjunction with the fresh perspective on possible structure to the space of all probabilistic epsilon machines provided by the concepts of “killing” and “collapsing”.

Understanding the Space of Epsilon Machines

When wishing to consider possible structure to the space of all probabilistic epsilon machines, it is helpful to be familiar with the development of epsilon machines in conjunction with the evolution of complexities studies. Over the course of the evolution of complexity studies, ideas like “complexity”, “pattern”, “order”, “structure”, “randomness”, “information”, and “memory” were formalized in an attempt to describe phenomena, their properties and their relationships in a way that allowed for a rigorous study of these subjects. Yet as more research in the field leads to a greater understanding of the nuances of these concepts, the definitions in place and the meaning(s) behind them continue to be revisited and reshaped. Thus, when it comes to a rigorous study of such subjects, it becomes imperative to be thoughtful in one’s understanding of the motivation behind definitions and in what respects these definitions can be applied meaningfully. Epsilon machines themselves are no exception.

In “Equivalence of History and Generator Epsilon Machines”, Travers and Crutchfield explain that the initial purpose of an epsilon machines was as a representation of a particular stationary stochastic process. Epsilon machines were formally defined as the minimal, unifilar presentations for stationary stochastic processes where the states of the epsilon machine were “equivalence classes of infinite past sequences that lead to the same predictions over future sequences”. This definition motivates questions involving epsilon machines as representations of specific processes, such as considering the robustness of statistical measurements taken from a given epsilon machine as discussed above.

Epsilon machines were then later defined as “irreducible, edge-label hidden Markov models with unifilar transitions and probabilistically distinct states”. This shift in perspective focuses on epsilon machines as a class of objects with specific properties and considers an arbitrary epsilon machine as the generator of a set of stationary stochastic processes. This definition motivates questions pertaining to a class of objects, and the structure of such a set.

Travers and Crutchfield go on to show that these two definitions are equivalent.

Much work has been done to better understand the structure of the space of epsilon machines, and many distinct approaches have been adopted. A successful approach was taken by Crutchfield, Johnson, Ellison, and McTague in “Enumerating Finitary Processes”, in which they construct an algorithm that enumerates all topological epsilon machines with **n** states and alphabet size **k**, where **n** and **k** are finite. Topological epsilon machines are a subset of probabilistic epsilon machines that are defined as having uniformly distributed transition probabilities on the edges exiting each state. This has significant consequences for the symmetries allowable in the structure of a topological epsilon machine, which are closely related to the processes of “killing” and “collapsing”.

As a probabilistic epsilon machine is defined by both its topology and symbol-label scheme as well as its transition probability distribution, a similar enumeration scheme for probabilistic epsilon machines is not possible due to the fact that, in general, for a given topology and symbol-labeling, there are an uncountable number of transition probability distributions an epsilon machine can take on.

What was discovered in this research is that probabilistic epsilon machines with certain topologies and symbol-labeling schemes cannot support certain transition probability distributions due to symmetries in the structure of the epsilon machine.

At these locations in the transition probability distribution parameter space of the epsilon machine, two phenomena occur that change the topology of the epsilon machine representation – “killing” and “collapsing”. “Killing” occurs when the transition probability on an edge of the epsilon machine goes to zero; the elimination of that edge from the epsilon machine representation can result in the removal of states from the epsilon machine representation as well as further “collapse”. “Collapsing” occurs when states of an epsilon machine no longer have distinct probability distributions over the future. In order for this representation of a stationary stochastic process to remain an epsilon machine representation of the process, these states must be associated thereby reducing the number of states in the epsilon machine representation. This paper lays out preliminary work in understanding “killing” and “collapsing” and offers promising results that these concepts will be beneficial in understanding the structure of the space of probabilistic epsilon machines.

Varying Transition Probabilities of Probabilistic Epsilon Machines

The transition probability distribution parameter space of a probabilistic epsilon machine with **n** states is an $(\mathbf{e_1} - 1)$-simplex $\bigotimes$ $(\mathbf{e_2} - 1)$-simplex $\bigotimes$ … $\bigotimes$ $(\mathbf{e_i} - 1)$-simplex $\bigotimes$ … $\bigotimes$ $(\mathbf{e_n} - 1)$-simplex, where $\mathbf{e_i}$ is the number of edges exiting state **i**. Note that $\mathbf{e_i}$ is bounded from above by the alphabet size. Each point in the transition probability distribution parameter space represents a probabilistic epsilon machine.

<u>Data</u>

Using code written in the CMPy server, we plotted the values of **hμ**, **Cμ**, and **E** over the transition probability distribution parameter space for epsilon machine representations of the Even Process, Golden Mean Process, Noisy Random Phase Slip Process, and Telescope Process.

*Refer to the Appendix to see the epsilon machine representations of these processes.

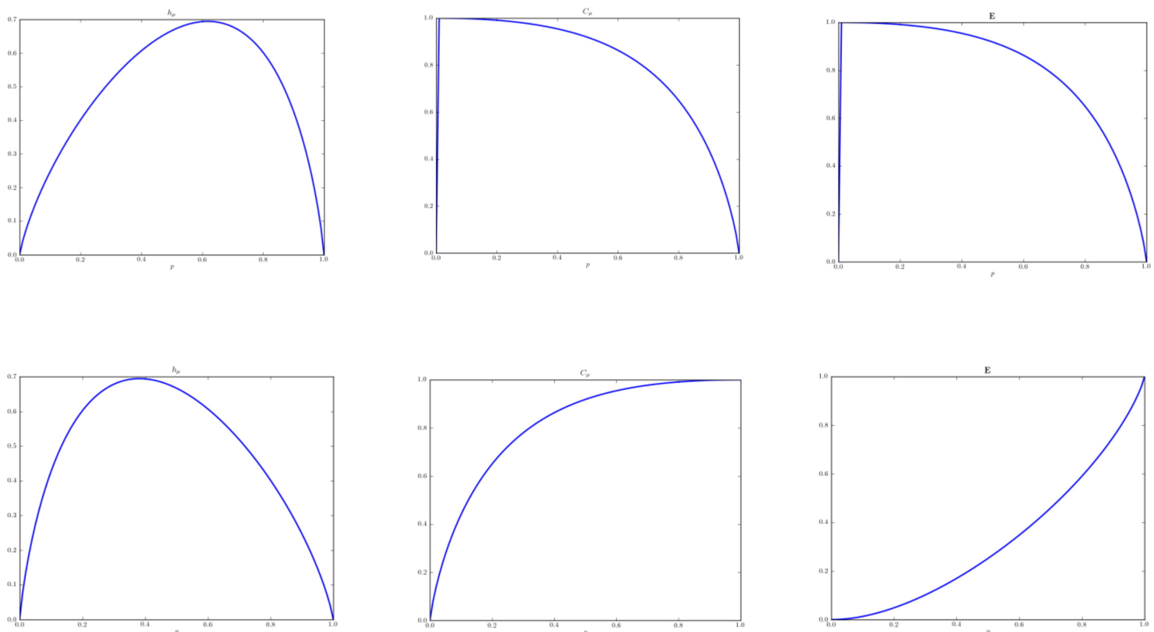
Below are the plots we generated: [insert key to colors –need to ask Ryan]

Fig.1

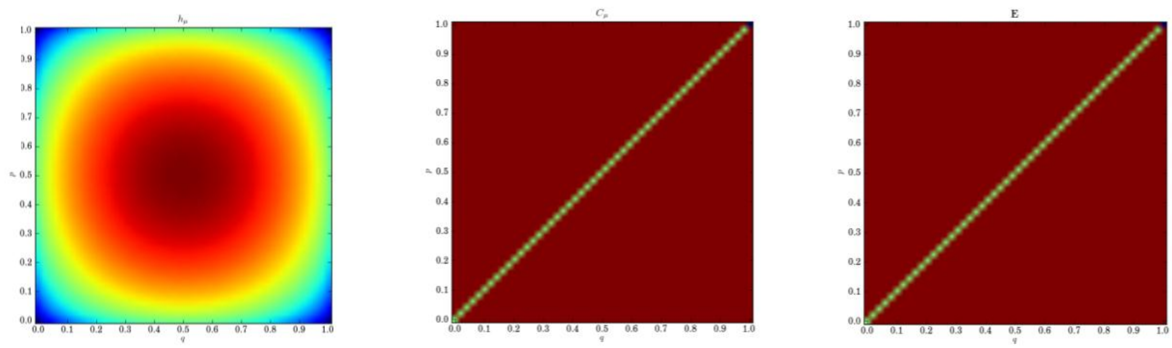
hμ: **Cμ:** **E:**

Even Process:

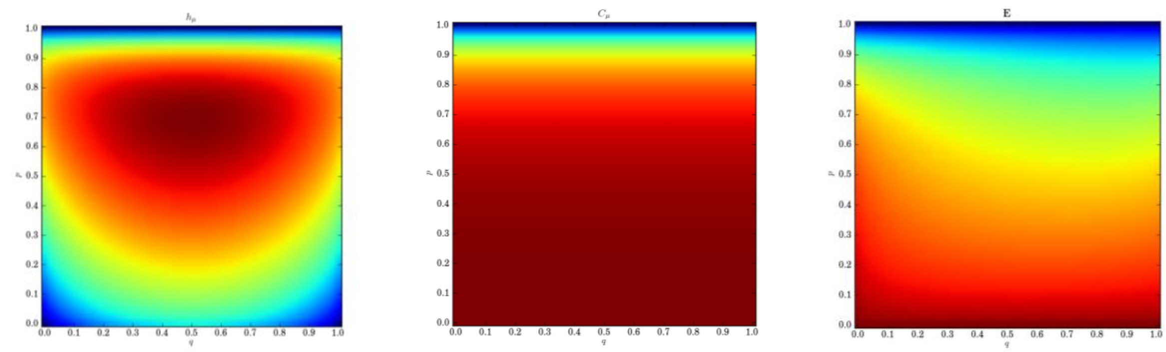
Golden Mean Process:



Telescope Process:



NRPS Process:



The alphabet size of the epsilon machine representations looked at in this research is 2, and the dimension of the transition probability distribution parameter space of these epsilon machines does not exceed 2. However, what was found encourages extending this research to consider these concepts in relation to higher dimensional transition probability distribution parameter spaces, as well, as a means of testing model robustness and understanding possible structure to the space of probabilistic epsilon machines.

<u>Analysis</u>

These results support the use of similar plots as a tool in determining model robustness. We can observe that in some regions of the parameter space, were our transition probabilities slightly off, we could get drastically different values for **hμ**, **Cμ**, and **E**. Even in regions of the parameter space where the values of **hμ**, **Cμ**, and **E** do not vary drastically, it is beneficial to know how they do vary in order to provide a bound on these quantities.

The diversity in how the entropy rate, statistical complexity, and excess entropy vary over the parameter space of transition probability distributions for these processes alone motivates using these measurements as a potential means of describing properties of probabilistic epsilon machines with isomorphic symbol-label schemes.

For instance, symmetries of these statistical and information-theoretic measurements within the transition probability distribution parameter space could reflect symmetries characteristic to the class of probabilistic epsilon machines with isomorphic symbol-label schemes over which the transition probabilities are being varied. This could be an interesting direction for future research, as epsilon machines are defined as a subset of a larger set of objects satisfying specific structural restrictions; thus, the structure of the space of epsilon machines is inherently connected to the structure of epsilon machines themselves.

Note also that the dimension of the transition probability distribution parameter space of a probabilistic epsilon machine of with **n** states is bound from above by the alphabet size, **k**, minus 1 times **n**. Looking at the set of all potential symbol-labeling schemes of all probabilistic epsilon machines achieving this topological “completeness”, as well as at the notion, and ramifications of, of alphabet size translation, could provide another perspective on the structure of the space of probabilistic epsilon machines.

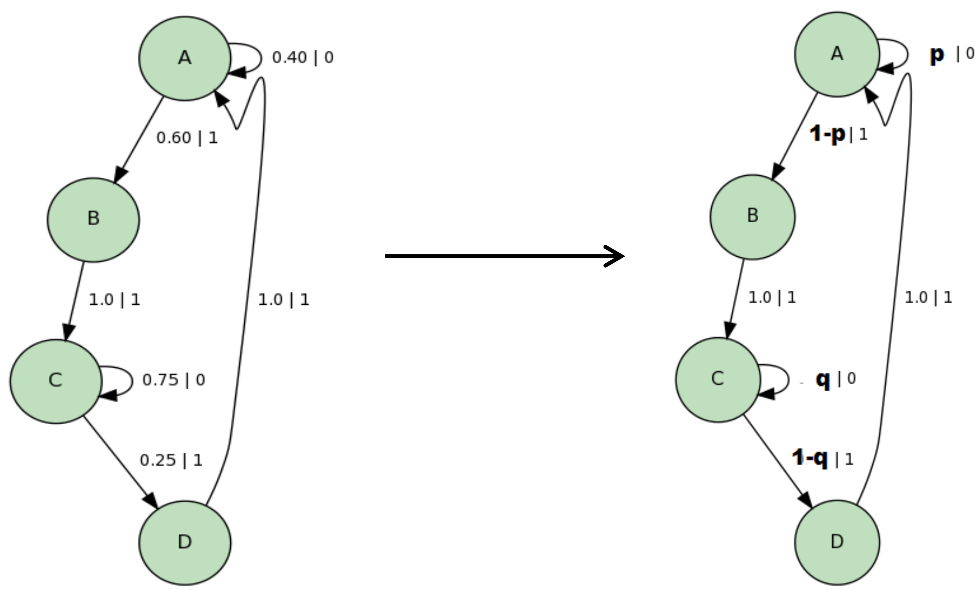
“Killing” and “Collapsing”: A Motivating Example

To motivate the concepts of “killing” and “collapsing”, we consider the following example.

<u>The Set Up</u>

Starting with the below probabilistic epsilon machine, we parameterize the transition probability distribution on the edges emanating from each state:

Fig.2:



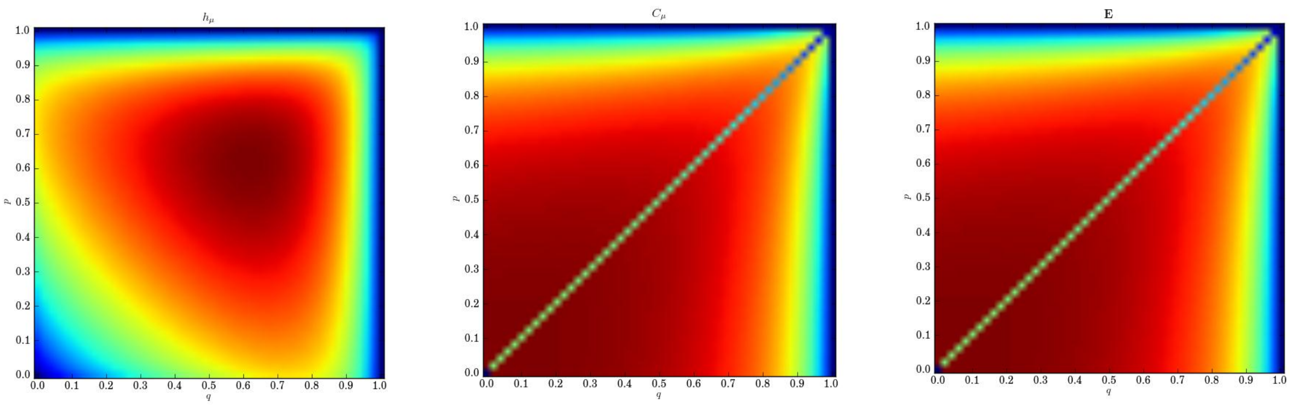
Probabilistic Epsilon Machine with Specific Transition Probabilities

Probabilistic Epsilon Machine with Parameterized Transition Probabilities

Plotting the values for the entropy rate, statistical complexity, and excess entropy over the transition probability distribution parameter space of this machine yields the following:

Fig.3:

hμ **Cμ** **E**

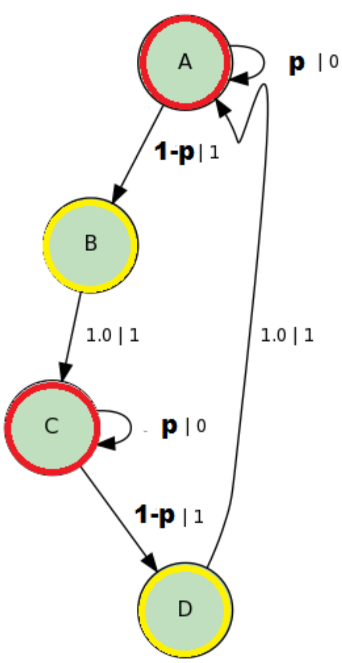


An analysis of these plots reveals that, unlike **hμ**, **Cμ** and **E** do not vary continuously near the diagonal where **p = q.** Another noteworthy feature is that the top and right boundary, where **q = 1** and **p = 1**, respectively, are uniform and identically zero for **hμ, Cμ**, and **E**. These features reveal topological consequences that occur as a result of varying the transition probability distribution.

<u>What’s Happening Where **p = q**: “Collapsing”</u>

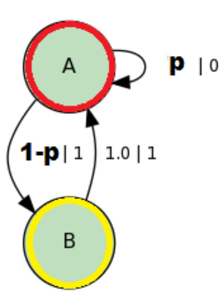
When **p** = **q**, the states **B** and **D**, as well as **A** and **C**, have the same probability distributions over futures.

Fig.4:



Machine Representation when **p = q**

Because states **B** and **D**, as well as **A** and **C**, have the same probability distributions over futures, the machine representation we have is no longer an epsilon machine. Yet this “non-epsilon” machine does represent a stationary stochastic process, therefore there is an epsilon machine representation for it. Our problem with this “non-epsilon” machine representation is that it is not minimal. We fix this by letting states **B** = **D** and **A** = **C**. Doing so “collapses” our machine into the familiar Even Process.



The Even Process

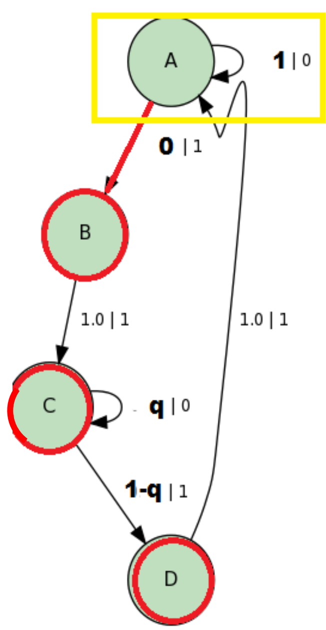
It is this process of associating states that alters the topology of the epsilon machine representation within the parameter space of a given epsilon machine which we will call “collapse”.

<u>What's Happening Where **q = 1** and **p = 1**: "Killing"</u>

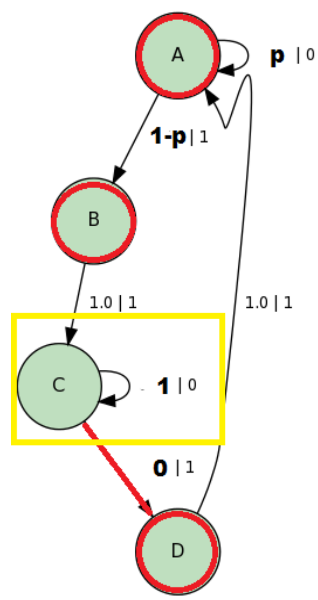
When **p** = **1**, the transition probability of the edge with symbol-label **0** from state **A** to state **B** goes to zero and we remove this edge from the epsilon machine representation. As a result of this, states **B**, **C**, and **D** become transient and are likewise removed from the epsilon machine representation. This leaves us with a single state epsilon machine that emits a single symbol, explaining why **hμ** = **Cμ** = **E** = **0** along the right boundary of the parameter space.

Similarly, when **q = 1** the transition probability of the edge with symbol-label **0** from state **C** to state **D** goes to zero and we remove that edge from the epsilon machine representation. Now, states **D**, **A**, and **B** become transient and are removed from the epsilon machine representation. This leaves us with a single state epsilon machine that emits a single symbol, explaining why **hμ** = **Cμ** = **E** = **0** along the top boundary of the parameter space.

Fig.5:



Machine Representation
when **p = 1**



Machine Representation
when **q = 1**

The process of removing an edge from the epsilon machine when the transition probability of it goes to **0** will be called "killing"; "killing" can result in the removal of states from the epsilon machine, as well as further "collapse".

Results

Using these concepts, we prove the following theorem and state the following conjecture. The potential implications of these to the study of the structure of the space of probabilistic epsilon machines are the objective of current research, and promising ideas will be introduced below.

Theorem

Theorem: Topological epsilon machines do not "collapse" in a region of the transition probability distribution parameter space that is not also a "killing**"** region.

Proof:

In order for states to "collapse", states "collapsing" must have the same probability distribution over the future. [and the same entering symbol-labeling with the same transition probabilities? Not needed for proof, but nonetheless important to think through if it is necessary for collapse in general]

Suppose we have a topological epsilon machine. Now suppose that this topological epsilon machine has a symbol-labeling scheme that supports a transition probability distribution that allows for some states to have the same probability distribution over the future [and same entering symbol-labeling with the same transition probabilities]. We wish to show this is impossible. We will do this by means of reaching a contradiction. Suppose the epsilon machine has a symbol-labeling scheme that allows for some states to also have the same probability distribution over the future. Then there must be the same number of edges exiting these states, and they must have the same symbol-labeling. By definition, the transition probabilities of topological epsilon machines are uniformly distributed. Since the number of edges exiting these states is the same, the exiting probabilities must also match up. This means that these potentially "collapsible" states have equivalent probability distributions for the future; thus, our machine is not minimal. We must conclude that this machine is not an epsilon machine. But, we supposed from the beginning that our machine was a topological epsilon machine. Having reached a contradiction, we can therefore say that we cannot have a topological epsilon machine that is "collapsible" without first "killing" off an edge. ▨

Conjecture

Conjecture: (The Converse) An epsilon machine with a symbol-labeling scheme that "collapses" only in a region of the transition probability distribution parameter space being transition parameter space that is also a "killing**"** region is isomorphic to a topological epsilon machine.

One way to approach this conjecture would be to consider the restrictions imposed on an epsilon machine by the condition of the “collapsing” region being contained within “killing” region of the transition probability distribution parameter space. The next step would be to demonstrate the equivalence between this set of restrictions and the enumeration scheme of topological epsilon machines found in “Enumerating Finitary Processes” as a pruning of a larger class of objects.

Implications

A potential implication of the above theorem and conjecture relates back to the work done by Crutchfield, Johnson, Ellison, and McTague in “Enumerating Finitary Processes” in which they enumerate all topological epsilon machines with **n** states and alphabet size **k**. Starting with an arbitrary probabilistic epsilon machine, we can look at the transition probability distribution parameter space and identify the region where “collapsing” occurs. Properties of this space as a submanifold of the transition probability distribution parameter space can alone be explored as a means of identifying structure to the space of all epsilon machines. Yet were the conjecture above proven true, the epsilon machine representations in the transition probability distribution parameter space where “collapsing” occurs are all topological epsilon machines, and thus can be identified according to the enumeration scheme presented in “Enumerating Finitary Processes”. Looking at the topological epsilon machines “nested” inside of a probabilistic epsilon machine and the structure of this “nesting” could possibly be used to classify probabilistic epsilon machines.

The perspective on the structure of the space of probabilistic epsilon machines provided by the concepts of “killing” and “collapsing” can also suggest other possible tests for model robustness. For instance, looking at the distance within the transition probability distribution parameter space between a probabilistic epsilon machine and the topological epsilon machines from which the probabilistic epsilon machine could have, after first “killing”, “collapsed” as well as the topological epsilon machines into which the probabilistic epsilon machine may “collapse” could provide some measure of model robustness.

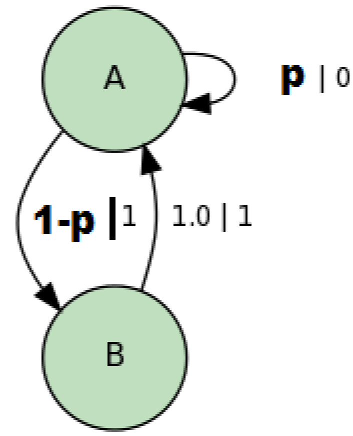
Conclusion

Though simple, the concepts of “killing” and “collapsing” seem to offer a powerful perspective when it comes to considering the structure of the space of probabilistic epsilon machines. The ability to better understand, and possibly make rigorous, the relationship between topological epsilon machines and probabilistic epsilon machines in relation to the structure of the space of all probabilistic epsilon machines via the concepts of “killing” and “collapsing” is just one application that demonstrates the wealth of knowledge to be gleaned by using these notions in tandem with previous research in the field.

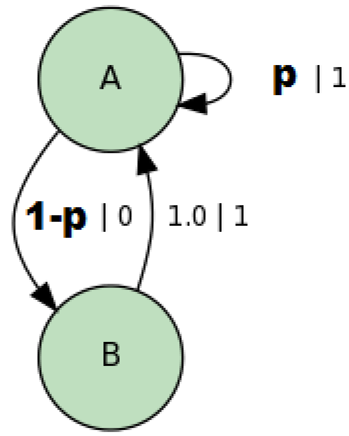
Appendix

Epsilon machines for the processes discussed in the paper are reproduced below:

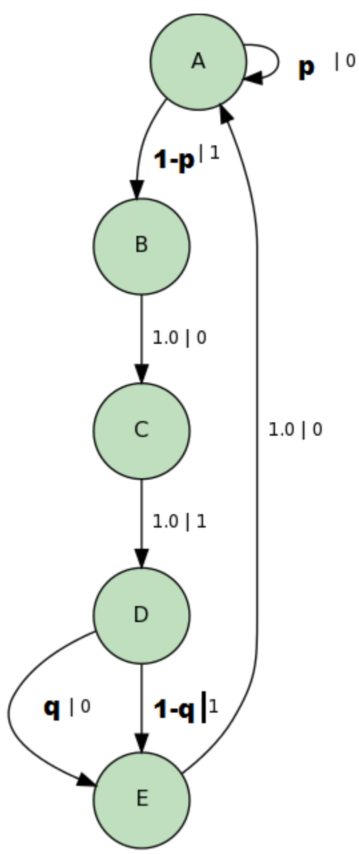
Fig.6:



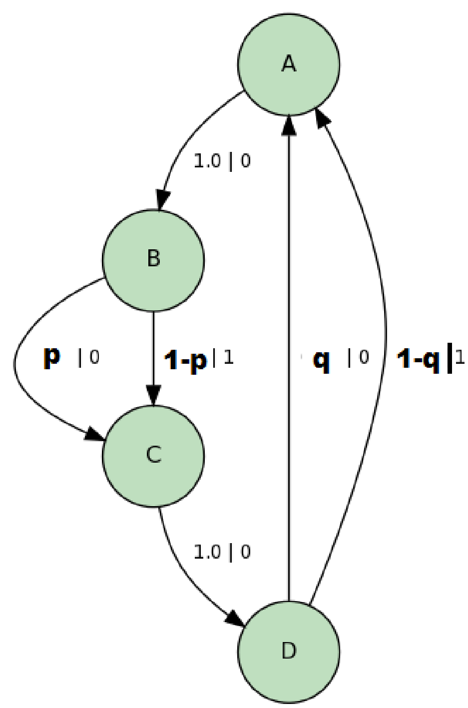
Even Process



Golden Mean Process



Noisy Random Phase Slip Process



Telescope Process

References

Crutchfield, J.P., Ellison, C.P., Johnson, B.D, McTague, C.S. 2011 “Enumerating Finitary Processes” **arXiv:1011.0036v2 [cs.FL]**

Crutchfield, J.P., Travers, N. F. 2011 “Equivalence of History and Generator Epsilon-Machines” **arXiv:1111.4500v1 [.PR]**

Crutchfield, J.P. “Lecture Notes” Natural Computation and Self-Organization Winter/Spring 2012