An Introduction to Statistical Complexity

MIR@W Statistical Complexity Day

University of Warwick

David P. Feldman

18 February 2008

College of the Atlantic

and

Santa Fe Institute

dave@hornacek.coa.edu

http://hornacek.coa.edu/dave/

Introduction

- This morning I will give a pedagogical introduction to a number of different measures of complexity and (un)predictability.
- This afternoon I will present some results that illustrate some interesting and fun properties of statistical complexity measures.
- I will also suggest some directions and opinionated guidelines for possible future work.
- My two lectures today are a very condensed version of a short course that I've developed for the Santa Fe Institute's Complex Systems Summer School in China, 2004–2007 and the ISC-PIF Complex Systems Summer School in Paris, 2007.
- These slides are at `hornacek.coa.edu/dave/Paris`. Please consult them for much more detail and many more references.

Outline

1. Why Complexity? Some context, history, and motivation.
2. Information Theoretic Measures of Unpredictability and Complexity
   (a) Entropy Rate
   (b) Excess Entropy
3. Computational Mechanics and Statistical Complexity

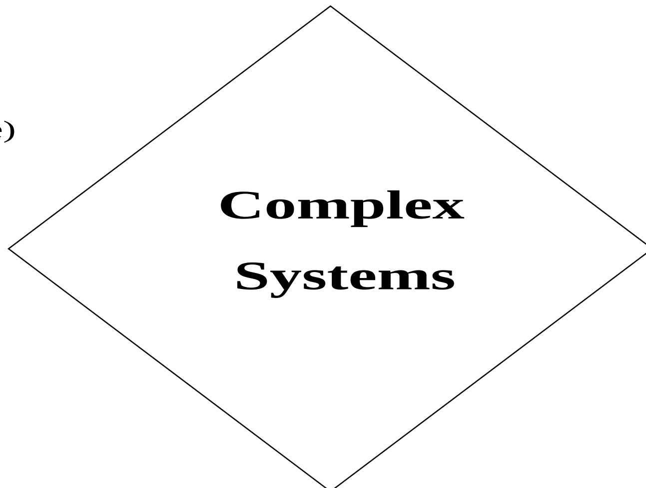
The next slide shows a highly schematic view of the universe of *complex systems* or *complexity science.*

Themes/General Principles??

Increasing Returns −−> "Power laws"
Stability Through Hierarchy
Stability through Diversity
Complexity Increases?
Exploitation vs. Exploration
And many more?

Topics/Models

Neural Networks (real & fake)
Spin Glasses
Evolution (real & fake)
Immune System
Gene Regulation
Pattern Formation
Soft Condensed Matter
Origins of Life
Origins of Civilization
Origin and Evolution of Language
Population Dynamics
And many, many, more...



Tools/Methods

Nonlinear Dynamics
Machine Learning
Cellular Automata
Symbolic Dynamics
Evolutionary Game Theory
Agent−Based Models
Information Theory
Stochastic Processes
Statistical Mechanics/RG
Networks
And many more ...

Foundations

Measures of Complexity
Representation and Detection of Organization
Computability, No Free Lunch Theorems
And many more...

Based on Fig. 1.1 from Shalizi, ”Methods and Techniques in Complex Systems Science: An Overview”, pp. 33-114 in Deisboeck and Kresh (eds.), Complex Systems Science in Biomedicine (New York: Springer-Verlag, 2006); `http://arxiv.org/abs/nlin.AO/0307015`

Comments on the Complex Systems Quadrangle

- The left and right hand corners of the quadrangle definitely exist.
- It is not clear to what extent the top of the quadrangle exists. Are there unifying principles? Loose similarities? No relationships at all?
- The bottom of the quadrangle exists, but may or may not be useful depending on one's interests.
- I'm not sure how valuable this figure is. Don't take it too seriously.
- Measures of complexity serve as a tool that can be used to understand model and real systems.
- I believe that measures of complexity also provide insight into fundamental questions about relationships between structure and randomness, and between the observer and the observed.

Complexity: Initial Thoughts

- The complexity of a phenomena is generally understood to be a measure of how difficult it to describe it.
- But, this clearly depends on the language or representation used for the description.
- It also depends on what features of the thing you're trying to describe.
- There are thus many different ways of measuring complexity. I will aim to discuss a bunch of these in my lectures.
- Some important, recurring questions concerning complexity measures:
  1. What does the measure tell us?
  2. Why might we want to know it?
  3. What representational assumptions are behind it?

Predictability, Unpredictability, and Complexity

- The world is an unpredictable place.
- There is predictability, too.
- But there is more to life than predictability and unpredictability.
- The world is patterned, structured, organized, complex.
- We have an intuitive sense that some things are more complex than others.
- Where does this complexity come from?
- Is this complexity real, or is it an illusion?
- How is complexity related to unpredictability (entropy)?
- What are patterns? How can they be discovered?

Information Theoretic View of Randomness and Structure

- Info theory was developed by Shannon in 1948.
- Information theory lets us ask and answer questions such as:
  1. How random is a sequence of measurements?
  2. How much memory is needed to store the outcome of measurements?
  3. How much information does one measurement tell us about another?
- Information theory provides a natural language for working with probabilities.
- Information theory is *not* a theory of semantics or meaning.

The Shannon entropy of a random variable X is given by:

$$H[X] \equiv - \sum_{x \in \mathcal{X}} \Pr(x) \log_2(\Pr(x)) \ . \tag{1}$$

Interpretations of Entropy

- $H[X]$ is *the* measure of uncertainty associated with the distribution of X.
- Requiring H to be a continuous function of the distribution, maximized by the uniform distribution, and independent of the manner in which subsets of events are grouped, uniquely determines H.
- $H[X]$ is the expectation value of the surprise, $-\log_2 \Pr(x)$.
- $H[X] \ \leq$ Average number of yes-no questions needed to guess the outcome of $X \ \leq \ H[X] + 1$.
- $H[X] \ \leq$ Average number of bits in optimal binary code for X $\leq \ H[X] + 1$.
- $H[X] = \lim N \to \infty \frac{1}{N} \times$ average length of optimal binary code of N copies of X.

Applying Information Theory to Stochastic Processes

- We now consider applying information theory to a long sequence of measurements.

$$\cdots 0011001001010110100110011101011 0\cdots$$

- In so doing, we will be led to two important quantities

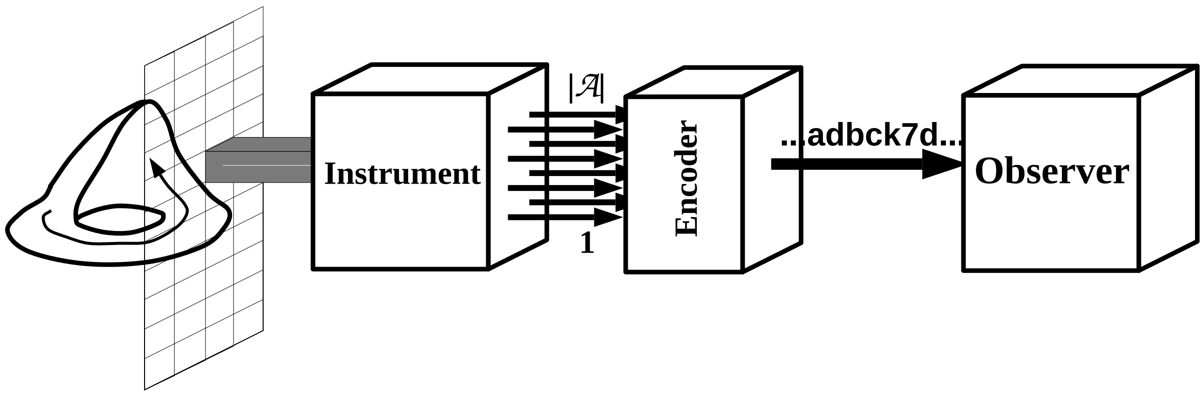
1. **Entropy Rate:** The irreducible randomness of the system.
2. **Excess Entropy:** A measure of the complexity of the sequence.

Context: Consider a long sequence of discrete random variables. These could be:

1. A long time series of measurements
2. A symbolic dynamical system
3. A one-dimensional statistical mechanical system

The Measurement Channel

- Can also picture this long sequence of symbols as resulting from a generalized measurement process:



- On the left is “nature”—some system’s state space.
- The act of measurement projects the states down to a lower dimension and discretizes them.
- The measurements may then be encoded (or corrupted by noise).
- They then reach the observer on the right.
- Figure source: Crutchfield, “Knowledge and Meaning ... Chaos and Complexity.” In Modeling Complex Systems. L. Lam and H. C. Morris, eds. Springer-Verlag, 1992: 66-10.

Stochastic Process Notation

- Random variables S_i, $S_i = s \in \mathcal{A}$.
- Infinite sequence of random variables: $\overleftrightarrow{S} = \ \ldots \ S_{-1}\, S_0\, S_1\, S_2\ \ldots$
- Block of L consecutive variables: $S^L \ = \ S_1, \ldots, S_L$.
- $\Pr(s_i, s_{i+1}, \ldots, s_{i+L-1}) \ = \ \Pr(s^L)$
- Assume translation invariance or stationarity:

$$\Pr(\,\mathrm{s_i}, \mathrm{s_{i+1}}, \cdots, \mathrm{s_{i+L-1}}\,) \ = \ \Pr(\,\mathrm{s_1}, \mathrm{s_2}, \cdots, \mathrm{s_L}\,)\ .$$

- Left half (“past”): $\overleftarrow{s} \equiv \cdots S_{-3}\, S_{-2}\, S_{-1}$
- Right half (“future”): $\overrightarrow{s} \equiv S_0\, S_1\, S_2 \cdots$

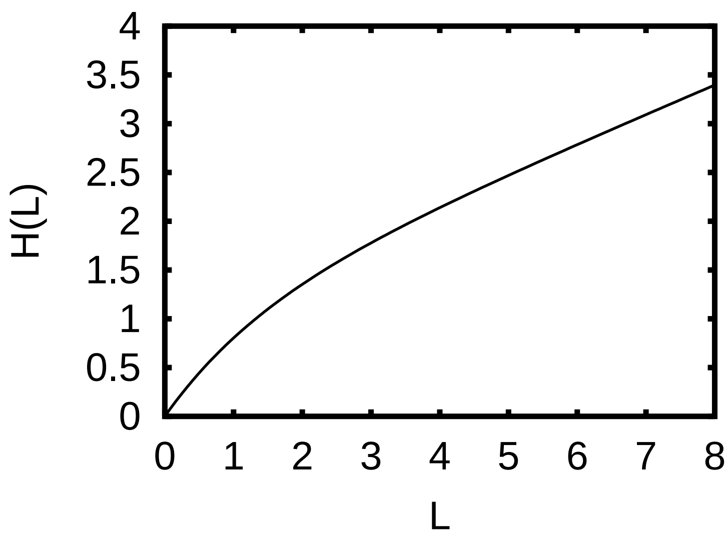
$$\cdots 11010100101101010101001001010010 \cdots$$

Entropy Growth

- Entropy of L-block:

$$H(L) \equiv -\sum_{s^L \in \mathcal{A}^L} \Pr(s^L) \log_2 \Pr(s^L) \ .$$

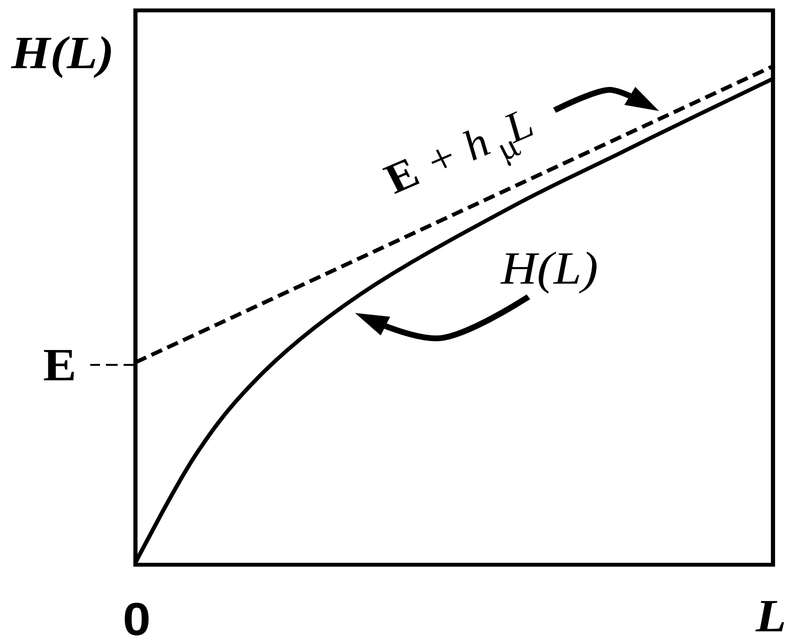
- $H(L) =$ average uncertainty about the outcome of L consecutive variables.



- $H(L)$ increases monotonically and asymptotes to a line
- We can learn a lot from the shape of $H(L)$.

Entropy Rate

- Let's first look at the slope of the line:



- Slope of $H(L)$: $h_\mu(L) \equiv H(L) - H(L-1)$
- Slope of the line to which $H(L)$ asymptotes is known as the *entropy rate:*

$$h_\mu = \lim_{L\to\infty} h_\mu(L).$$

Entropy Rate, continued

- Slope of the line to which $H(L)$ asymptotes is known as the *entropy rate:*

$$h_\mu = \lim_{L\to\infty} h_\mu(L).$$

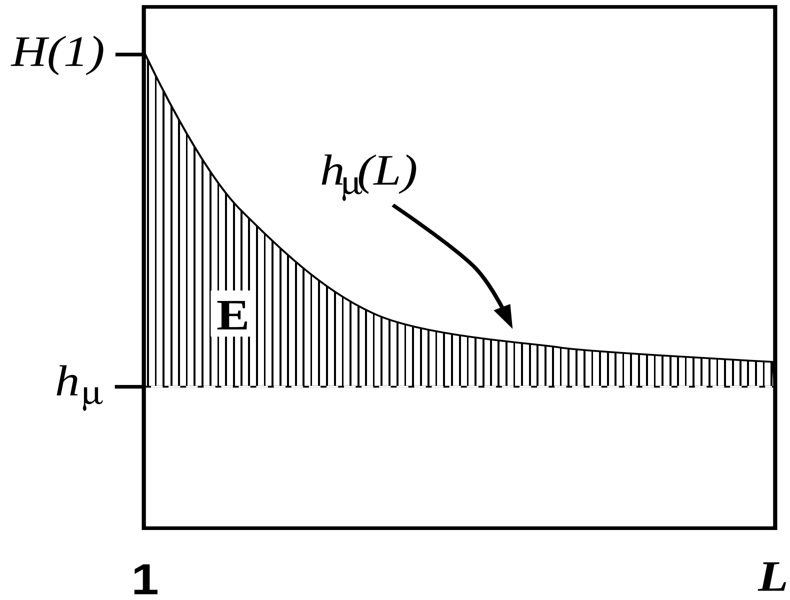
- $h_\mu(L) = H[S_L|S_1 S_1 \ldots S_{L-1}]$
- I.e., $h_\mu(L)$ is the average uncertainty of the next symbol, given that the previous L symbols have been observed.

Interpretations of Entropy Rate

- Uncertainty per symbol.
- Irreducible randomness: the randomness that persists even after accounting for correlations over arbitrarily large blocks of variables.
- The randomness that cannot be “explained away”.
- Entropy rate is also known as the Entropy Density or the Metric Entropy.
- $h_\mu =$ Lyapunov exponent for many classes of 1D maps.
- The entropy rate may also be written: $h_\mu = \lim_{L\to\infty} \frac{H(L)}{L}$.
- h_μ is equivalent to thermodynamic entropy.
- These limits exist for all stationary processes.

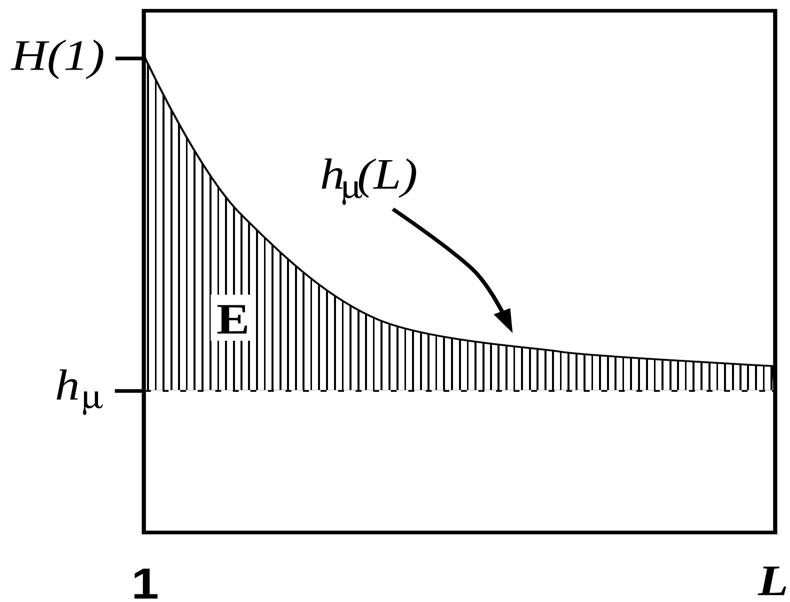
How does $h_\mu(L)$ approach h_μ?

- For finite L , $h_\mu(L) \geq h_\mu$. Thus, the system appears more random than it is.



- We can learn about the complexity of the system by looking at *how* the entropy density converges to h_μ.

The Excess Entropy



- The **excess entropy** captures the nature of the convergence and is defined as the shaded area above:

$$\mathbf{E} \equiv \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu] \ .$$

- $\mathbf{E}$ is thus the total amount of randomness that is "explained away" by considering larger blocks of variables.

Excess Entropy: Other expressions and interpretations

Mutual information

- One can show that $\mathbf{E}$ is equal to the mutual information between the "past" and the "future":

$$\mathbf{E} = I(\overleftarrow{S}; \overrightarrow{S}) \equiv \sum_{\{\overleftrightarrow{s}\}} \Pr(\overleftrightarrow{s}) \log_2 \left[\frac{\Pr(\overleftrightarrow{s})}{\Pr(\overleftarrow{s})\Pr(\overrightarrow{s})} \right] .$$

- The Mutual Information $I[X;Y]$ is defined as the reduction in uncertainty about one variable given the outcome of the other:

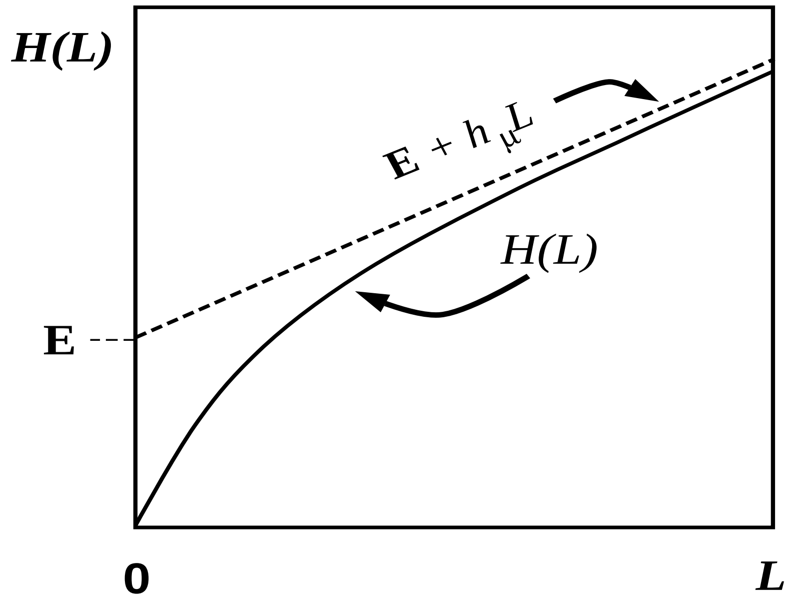
$$I[X;Y] = H[X] - H[X|Y] .$$

- $\mathbf{E}$ is thus the amount one half "remembers" about the other, the reduction in uncertainty about the future given knowledge of the past.
- Equivalently, $\mathbf{E}$ is the "cost of amnesia:" how much more random the future appears if all historical information is suddenly lost.

Excess Entropy: Other expressions and interpretations

Geometric View

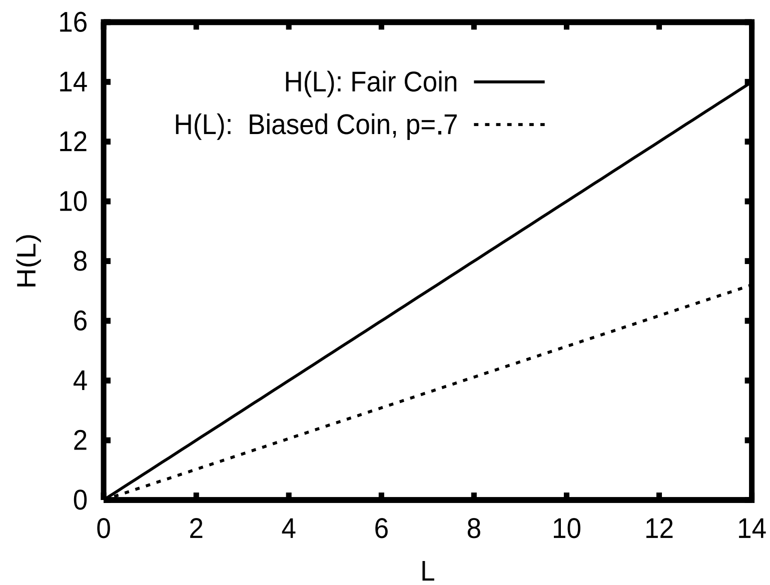
- $\mathbf{E}$ is the y-intercept of the straight line to which $H(L)$ asymptotes.
- $\mathbf{E} = \lim_{L \to \infty} [H(L) - h_\mu L]$.



Excess Entropy Summary

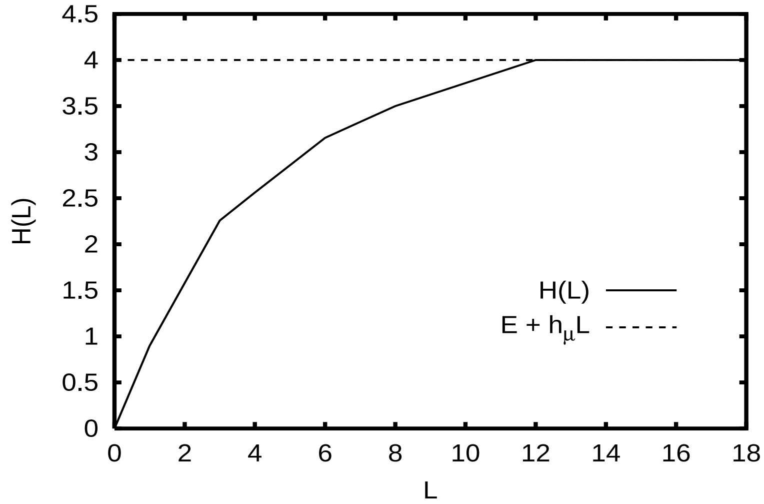
- Is a structural property of the system — measures a feature complementary to entropy.
- Measures memory or spatial structure.
- Lower bound for statistical complexity, minimum amount of information needed for minimal stochastic model of system

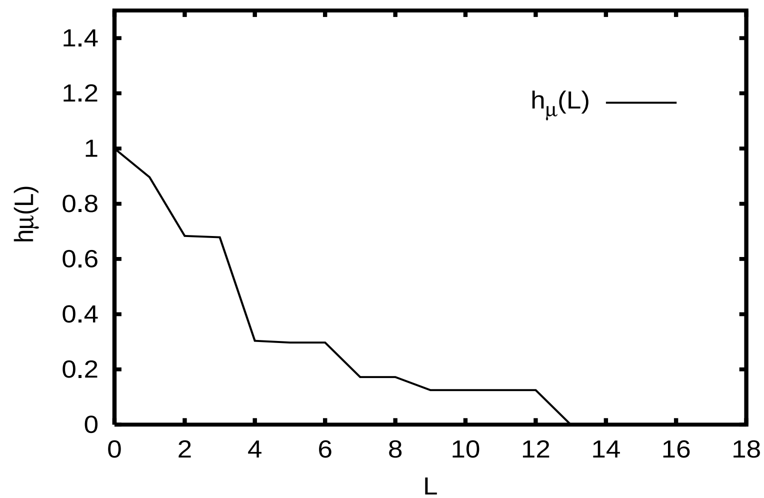
Example I: Fair Coin



- For fair coin, $h_\mu = 1$.
- For the biased coin, $h_\mu \approx 0.8831$.
- For both coins, $\mathbf{E} = 0$.
- Note that two systems with different entropy rates have the same excess entropy.

Example II: Periodic Sequence





- Sequence: $\ldots 1010111011101110 \ldots$

Example II, continued

- Sequence: $\ldots 1010111011101110 \ldots$
- $h_\mu = 0$; the sequence is perfectly predictable.
- $\mathbf{E} = \log_2 16 = 4$: four bits of phase information
- For any period-p sequence, $h_\mu = 0$ and $\mathbf{E} = \log_2 p$.

For many more examples, see Crutchfield and Feldman, *Chaos*, 15: 25-54, 2003.

For more than you probably ever wanted to know about periodic sequences, see Feldman and Crutchfield, Synchronizing to Periodicity: The Transient Information and Synchronization Time of Periodic Sequences. *Advances in Complex Systems*. **7**(3-4): 329-355, 2004.

Excess Entropy: Notes on Terminology

All of the following terms refer to essentially the same quantity.

- **Excess Entropy:** Crutchfield, Packard, Feldman
- **Stored Information:** Shaw
- **Effective Measure Complexity:** Grassberger, Lindgren, Nordahl
- **Reduced (Rényi) Information:** Szépfalusy, Györgyi, Csordás
- **Complexity:** Li, Arnold
- **Predictive Information:** Nemenman, Bialek, Tishby

Excess Entropy: Selected References and Applications

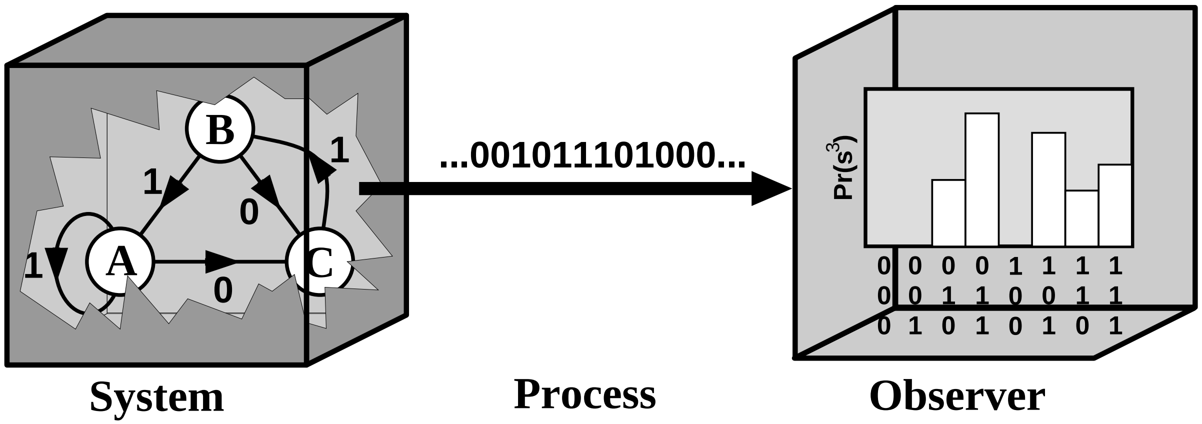
- Crutchfield and Packard, *Intl. J. Theo. Phys*, 21:433-466. (1982); *Physica D*, 7:201-223, 1983. [Dynamical systems]
- Shaw, "The Dripping Faucet ..., " Aerial Press, 1984. [A dripping faucet]
- Grassberger, *Intl. J. Theo. Phys*, 25:907-938, 1986. [Cellular automata (CAs), dynamical systems]
- Szépfalusy and Györgyi, *Phys. Rev. A*, 33:2852-2855, 1986. [Dynamical systems]
- Lindgren and Nordahl, *Complex Systems*, 2:409-440. (1988). [CAs, dynamical systems]
- Csordás and Szépfalusy, *Phys. Rev. A*, 39:4767-4777. 1989. [Dynamical Systems]
- Li, *Complex Systems*, 5:381-399, 1991.
- Freund, Ebeling, and Rateitschak, *Phys. Rev. E*, 54:5561-5566, 1996.
- Feldman and Crutchfield, SFI:98-04-026, 1998. Crutchfield and Feldman, *Phys. Rev. E* 55:R1239-42. 1997. [One-dimensional Ising models]

Excess Entropy: Selected References and Applications, continued

- Feldman and Crutchfield. *Physical Review E*, 67:051104. 2003. [Two-dimensional Ising models]
- Feixas, et al, *Eurographics*, Computer Graphics Forum, 18(3):95-106, 1999. [Image processing]
- Ebeling. *Physica D*, 1090:42-52. 1997. [Dynamical systems, written texts, music]
- Bialek, et al, *Neur. Comp.*, 13:2409-2463. 2001. [Long-range 1D Ising models, machine learning]

Estimating Probabilities

- $\mathbf{E}$ and h_μ can be estimated empirically by observing a process.



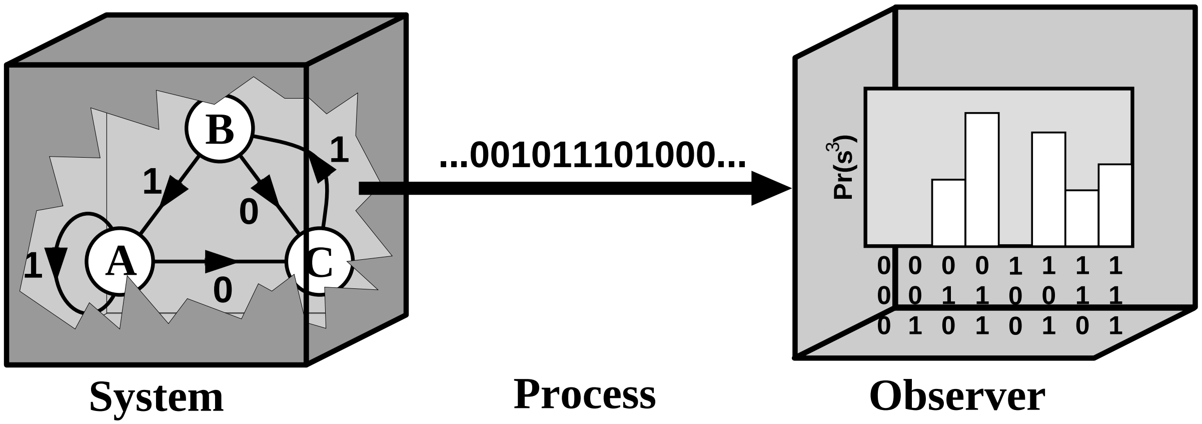
- One simply forms histograms of occurrences of particular sequences and uses these to estimate $\mathrm{Pr}(s^L)$, from which $\mathbf{E}$ and h_μ may be readily calculated.

However, this will lead to a biased under-estimate for h_μ. For more sophisticated and accurate ways of inferring h_μ, see, e.g.,

- Schürmann and Grassberger. Chaos 6:414-427. 1996.
- Nemenman. `http://arXiv.org/physics/0207009`. 2002.

A look ahead

- Note that the observer sees measurement symbols: 0's and 1's.



- It doesn't see inside the "black box" of the system.
- In particular, it doesn't see the internal, hidden states of the system, A, B, and C.
- Is there a way an observer can infer these hidden states?
- What is the meaning of *state*?

An Introduction to Computational Mechanics

1. Computational Mechanics provides another way of measuring an object's complexity or regularities.
2. Unlike the excess entropy, computational mechanics makes use of the models of formal computation to provide a direct, structural accounting of a system's intrinsic information processing.
3. Computational Mechanics lets us see how a system stores, transmits, and manipulates information.

Context:

- As before, we have a long sequence of symbols, $s_1, s_2, s_3, \cdots$, from a binary alphabet. Assume a stationary probability distribution over the sequence.

An Initial Example: The Prediction Game

- Your task is to observe a sequence, and then come up with a way of predicting, as best you can, subsequent values of the sequence.
- The sequence might have non-zero entropy rate, so perfect prediction might be impossible.
- We will begin by focusing at some length on the following example:

$$\ldots 1011111010111011101 0111 \ldots$$

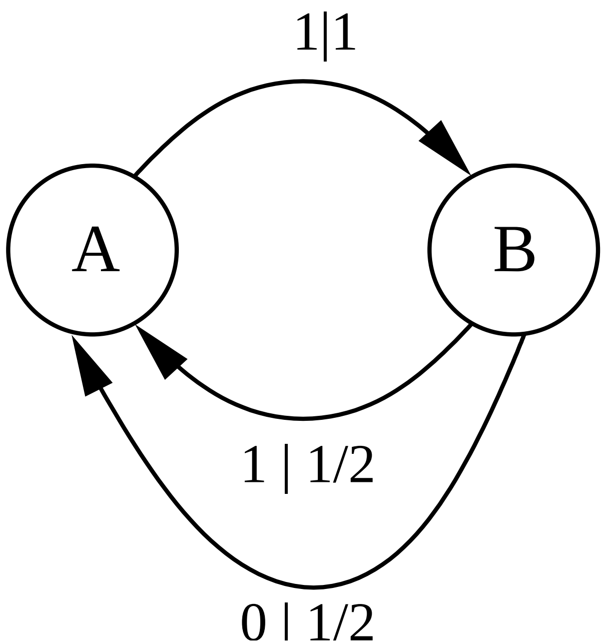
Discovery!

$$\ldots 1011111010111011101 0111 \ldots$$

- After some squinting, you will probably notice that every other symbol is 1. The other symbols are 0 or 1 with equal probability.
- You discovered a pattern: a regularity.
- Note that this pattern is stochastic.
- Note that you did not *recognize* the pattern.
- Recognition entails searching for a match to a pre-determined set of patterns or templates.
- Discovery means finding something new: something not necessarily seen before.
- How can we represent this regularity mathematically, and can we program a computer to do pattern discovery?

Initial example, continued

- The machine that can reproduce this sequence is:



- From state $\mathbf{A}$, one sees a 1 with probability 1.
- From sate $\mathbf{B}$, one sees a 1 with probability $1/2$, and a 0 with probability $1/2$.
- This is a stochastic generalization of a finite state machine.
- Note that it is still *deterministic* in the sense that the output symbol (0 or 1) determines the next state ($\mathbf{A}$ or $\mathbf{B}$).

Initial Example: Why Two States?

- Why are only two states necessary? And what exactly do we mean by “state”?
- There are many particular observed sequences which give one equivalent information about the future sequences
- For example, if you see 1010, or 1110 or simply 0, in all cases you know with certainty that a 1 is next.
- The idea is that it only makes sense to distinguish between historical sequences that give rise to different predictive information.
- There will usually be many sequences that give the same predictive information. Group these sequences together into a **state**.
- These states are known as **causal states**. I will formalize this notion of state below.

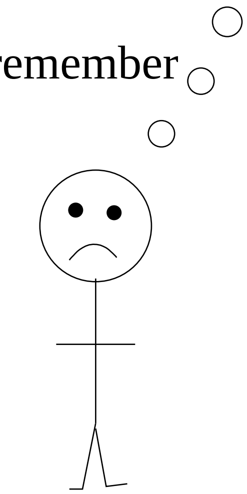
What do you Need to Remember in Order to Predict?

Space of all possible pasts.

01111 0101 11011
011 10111 010
1111 1
0111 011 1111 01110
10101
0 01111
110 111
11110 01 1011 1110
1010 01011 111111
11 101 10
11111 11101 1101
110111 010111

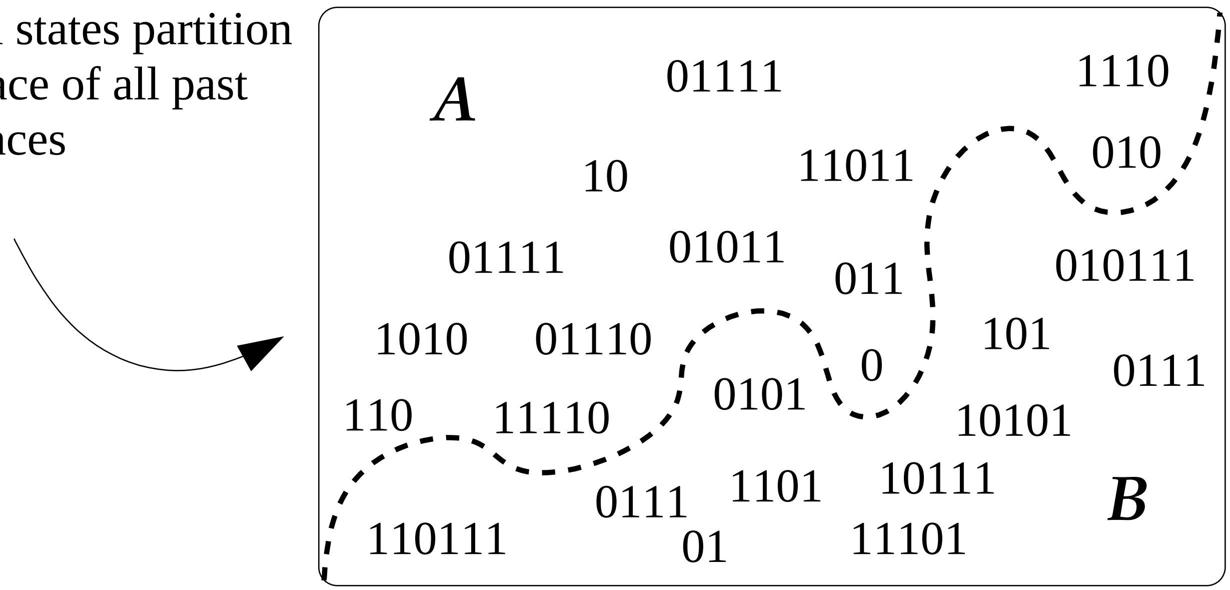
Do I really have to remember all this??

My memory isn't good enough.



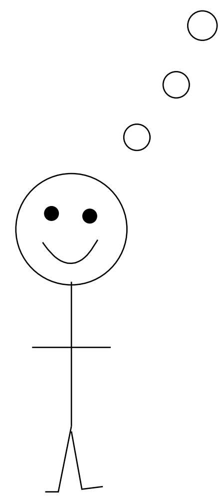
One Only Needs to Remember the Causal States.

Causal states partition the space of all past sequences



This is better!

I only need to remember the causal state, A or B.



How Might We Find Causal States?

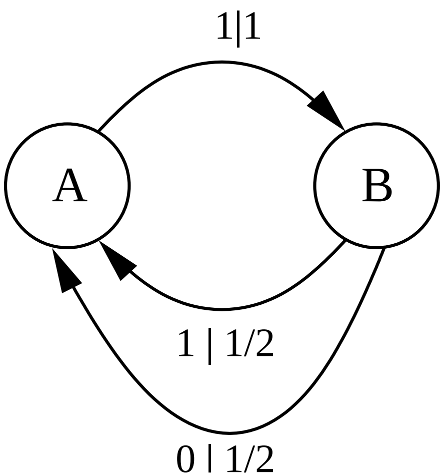
- How much of the left half $\overleftarrow{S}$ is needed to predict the right half $\overrightarrow{S}$?
- Only need to distinguish between $\overleftarrow{S}$'s that give rise to different states of knowledge about $\overrightarrow{S}$.
- Two $\overleftarrow{S}$'s that give rise to the same state of knowledge are equivalent:

$$\overleftarrow{S}_i \sim \overleftarrow{S}_j \ \text{ iff} \Pr(\overrightarrow{S} \mid \overleftarrow{s}_i) = \Pr(\overrightarrow{S} \mid \overleftarrow{s}_j) \ .$$

- Equivalence classes induced by $\sim$ are **Causal States**, minimal sets of aggregate variables necessary for optimal prediction of $\overrightarrow{S}$.
- For example, $\Pr(\overrightarrow{S} |0) = \Pr(\overrightarrow{S} |1011)$. Hence, 0 and 1011 are equivalent under $\sim$.
- This means that the probability over the futures $\overrightarrow{S}$ is the same if you've seen 0 or 1011.

ϵ-Machines

- The causal states together with the probability of transitions between causal states are an ϵ**-machine**, a minimal model capable of statistically reproducing the original configuration.
- The ϵ-machine tells us **how the system computes**.
- The "ϵ" reminds us that the measurement symbols upon which the machine is formed may be distorted via noise or the discretization process.



- Note: In this example $h_\mu = 1$.

Distribution over Causal States

- Transitions between causal states are Markovian.
- Thus, the stationary (or asymptotic) distribution $p \equiv \Pr(\sigma)$ over the causal states is the left eigenvector of the transition matrix T:

$$pT = p \; . \tag{2}$$

- Normalize p so that $\sum_\alpha p_\alpha = 1$.
- For this example,

$$p = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \; . \tag{3}$$

- I.e., the ϵ-machine spends an equal amount of time in states $\mathbf{A}$ and $\mathbf{B}$.

Statistical Complexity

- The *statistical complexity* is defined as the Shannon entropy of the asymptotic distribution of the causal states:

$$C_\mu \equiv -\sum_\alpha p_\alpha \log_2 p_\alpha \ . \tag{4}$$

- To perform optimal prediction of the system one needs only to remember the causal states.
- The statistical complexity thus measures the minimum amount of memory needed to perform optimal prediction.
- The statistical complexity is a measure of the pattern or structure or regularity present in the system.
- For our example, $C_\mu = 1$.

Some Important Properties of ϵ-machines

- (For proofs, see Shalizi and Crutchfield. *J. Statistical Physics*. **104**:819. 2001.)
- The causal states are a *sufficient statistic*:

$$I[\overrightarrow{S}; \overleftarrow{S}] = I[\overrightarrow{S}; \sigma] \ . \tag{5}$$

  I.e., all the information about the future is contained in the causal states.

- The causal states are minimal.
- The causal states are unique up to trivial relabeling.
- The causal states form a Markov process.
- The ϵ-machine is a semi-group.

Statistical Complexity vs. Excess Entropy

- Both the statistical complexity C_μ and the excess entropy $\mathbf{E}$ are measures of complexity or structure or pattern or organization. However, they are not the same.
- $C_\mu =$ the minimal amount of memory needed to optimally predict the process.
- $\mathbf{E} =$ the amount of information the past carries about the future.

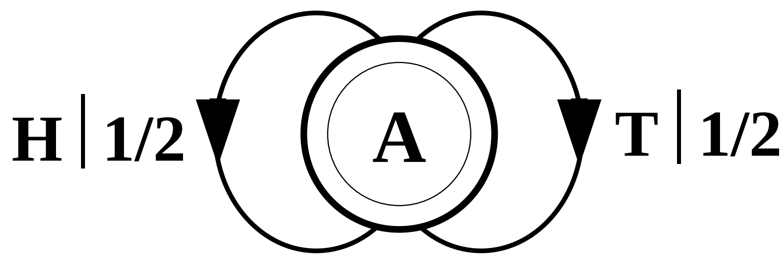
$$C_\mu \geq \mathbf{E} \,. \tag{6}$$

$$\text{Memory needed for model} \geq \text{Memory of the process itself} \,. \tag{7}$$

- $\mathbf{E}$ is time reversal invariant; C_μ is not.

Example I

Fair Coin:

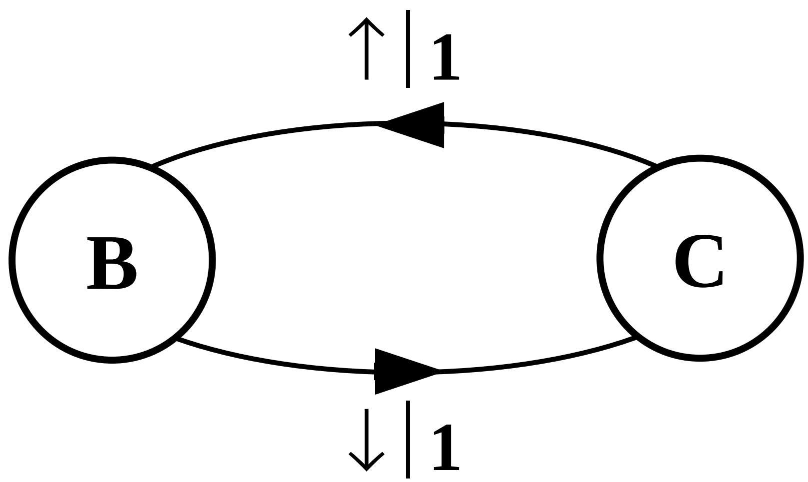


$$\cdots \mathrm{HHTHTHTTTHTHTHTTHTHH} \cdots$$

Entropy rate $h_\mu = 1$, Statistical Complexity $C_\mu = 0$.

Example II

Period 2 Pattern:

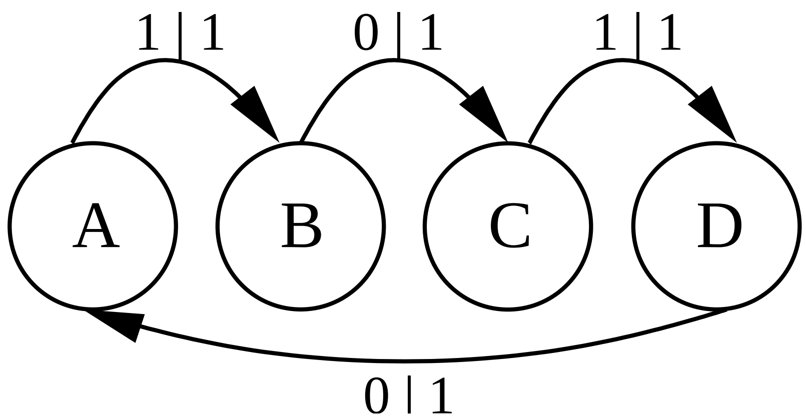


$$\cdots \uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow \cdots$$

Entropy rate $h_\mu = 0$, Statistical complexity $C_\mu = 1$.

A non-minimal example

Consider this machine for a period 2 sequence:



- States A and C are identical—they represent the same state of information about the future.
- So A and C should be merged to make one causal state.
- The same holds for B and D.
- The process of forming equivalence classes described on previous slides ensure that ϵ-machines are minimal.

Algorithms for Inferring ϵ-machines

There are two basic approaches

1. **Merge**
   - Initially distinguish between different histories. Then *merge* states that give rise to the same future distribution. I.e., merge states that are equivalent under $\sim$.
   - See Hanson, *PhD Thesis*, University of California, Berkeley, 1993.
2. **Split:**
   - Start with one state. This is equivalent to assuming a history of length zero. I.e., an IID process.
   - Add a symbol to history length. Split each state only if doing so increases predictability.
   - Repeat.

CSSR

- Shalizi and Shalizi(Klinkner) have implemented a state-splitting algorithm known as CSSR. (Causal State Splitting Algorithm)
- See Shalizi and Shalizi pp. 504–511 of Max Chickering and Joseph Halpern (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference*, `http://arxiv.org/abs/cs.LG/0406011`.
- See also Shalizi, Shalizi, and Crutchfield. `http://arxiv.org/abs/cs.LG/0210025`. 2002.
- CSSR source code is available at `http://bactra.org/CSSR`.
- CSSR has been applied to: crystallography, geomagnetic fluctuations, natural languages, anomaly detection, natural languages, and more.

Computational Mechanics References and Applications

Almost all of the papers below can be found online either on arXiv.org or with a little bit of searching.

- Crutchfield and Young, *Phys. Rev. Lett*, 63:105-108, 1989
- Crutchfield and Young, in *Complexity, Entropy and the Physics of Information*, Addison-Wesley, 1990. [Detailed analysis of Logistic and Tent maps]
- Crutchfield, *Physica D*, 75:11-54, 1994. [Long article, good review section, many different examples. A good place to start.]
- Shalizi and Crutchfield. *J. Statistical Physics*. **104**:819. 2001. [Mathematical foundations of causal states. Careful proofs of optimality and minimality.]

Applications and Extensions of Causal States

- Hanson, *PhD Thesis*, University of California, Berkeley, 1993. [Cellular Automata]
- Hanson and Crutchfield, *Physica D*, 103:169-189, 1997. [Cellular Automata]
- Upper, *PhD Thesis*, University of California, Berkeley, 1997. [Hidden Markov Models]
- Delgado and Solé, *Phys. Rev. E*, 55:2338-2344, 1997. [Coupled Map Lattices]
- Witt, Neiman and Kurths, *Phys. Rev. E*, 55:5050-5059, 1997. [Stochastic resonance]
- Goncavales, et. al., *Physica A*, 257, 385-389. 1998. [Dripping faucets]
- Feldman and Crutchfield, SFI:98-04-026, 1998. [One-dimensional Ising models. Includes lengthy review, calculations of excess entropy, and comparisons to statistical mechanical quantities.]
- Varn, et al. *Physical Review B*. **66**:156. 2002. [Layered Solids]
- Clarke, et al. *Physical Review E*. **67**:016203. 2003 [Geomagnetism]
- Palmer, et al. *Advances in complex systems*. 1:1-16. 2001. [Climate modeling, ϵ-machines inferred from empirical data.]
- Shalizi, Discrete Mathematics and Theoretical Computer Science, AB(DMCS) (2003): 11-30. [Dynamical systems on random networks]

Applications and Extensions of Causal States, Continued

- Görnerup and Crutchfield. SFI 04-06-020. [Self-assembling evolutionary systems]
- Ray. *Signal Processing*. **84**:1114. 2004.
- Shalizi, et al. *Physical Review Letters*. **93**:118701. 2004. [Cellular automata in more than one dimension]
- Padro and Padro, in *Proceedings of the Fifth International Workshop on Finite-State Methods and Natural Language Processing.* 2005.
- Young, et al. *Physical Review Letters.* **94**:098701. 2005. [Two-dimensional brain slices. Applications to Alzheimer's disease.]
- Park, et al. *Physica A*. **379**:179. 2007. [Financial time series. Stock market.]
- Klinkner, et al. `arXiv:q-bio/0506009v2`. [Shared information in neural networks.]
- Shalizi, et al. *Phys. Rev.E*. **73**: 036104. 2006. [2D cellular automata. Automatic order-parameter finding!]

Computational Mechanics Conclusions:

Questions:

- What are patterns and how can we discover them?
- What does it mean to say a system is organized?

Summary:

- Computation theory classifies sets of sequences by considering how difficult it is to recognize them.
- Causal states and ϵ-machines adapt computation theory for use in a probabilistic setting.
- The ϵ-machine provides an answer to the question: What patterns are present in a system?
- The ϵ-machine can be inferred directly from observed data.
- The ϵ-machine reconstruction pattern can discover patterns—even patterns that we haven't seen before.

The Objective Subjectivity of Complexity

MIR@W Statistical Complexity Day

University of Warwick

David P. Feldman

18 February 2008

College of the Atlantic

and

Santa Fe Institute

dave@hornacek.coa.edu

http://hornacek.coa.edu/dave/

Outline

1. Four examples illustrating the subjectivity or contextuality of complexity.

2. Exploring the the relationship between complexity and entropy.

3. Some thoughts on possible futures for complexity measures.

Thoughts on the Subjectivity of Complexity

- There is not a general, all-purpose, objective measure of complexity.
- Objective knowledge is, in a sense, knowledge without a knower.
- Subjective knowledge depends on the knower.
- Complexity, at least as I've been using the term, is a measure of the difficulty of describing or modeling a system.
- This will depend on who is doing the observing and what assumptions they make.
- Depending on the observer a system may appear more or less complex.
- Entropy and complexity are often related in interesting ways.
- I'll illustrate this with four examples.

Example I: Disorder as the Price of Ignorance

- Let us suppose that an observer seeks to estimate the entropy rate.
- To do so, it considers statistics over sequences of length L and then estimates h_μ using an estimator that assumes $\mathbf{E} = 0$.
- Call this estimated entropy ${h_\mu}'(L)$. Then, the difference between the estimate and the true h_μ is (Prop. 13, Crutchfield and Feldman, 2003):

$$h'_\mu(L) - h_\mu \; = \; \frac{\mathbf{E}}{L} \; .$$

- In words: The system appears more random than it really is by an amount that is directly proportional to the the complexity $\mathbf{E}$.
- In other words: regularities ($\mathbf{E}$) that are missed are converted into apparent randomness ($h'_\mu(L) - h_\mu$).
- Crutchfield and Feldman, “Regularities Unseen, Randomness Observed.” *Chaos*. 15:23-54. 2003.

Example II: Effects of Bad Discretization

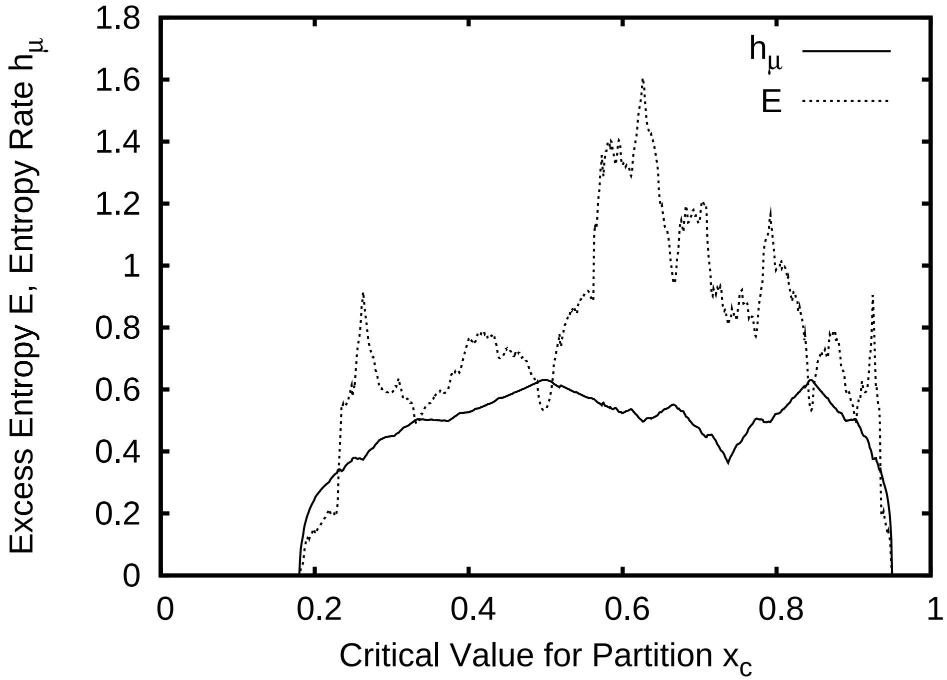
- Iterate the logistic equation: $x_{n+1} = f(x_n)$, where $f(x) = rx(1-x)$.
- Result is a sequence of numbers. E.g., $0.445, 0.894, 0.22, 0.344, \ldots$.
- Generate symbol sequence via:

$$s_i = \begin{cases} 0 & x \leq x_c \\ \\ 1 & x > x_c \end{cases} .$$

- For many values of r this system is chaotic.
- It is well-known that if $x_c = 0.5$, then the entropy of the symbol sequence is equal to the entropy of the original sequence of numbers.
- Moreover, it is well known that h_μ is maximized for $x_c = 0.5$.

Example II: Effects of Bad Discretization (continued)

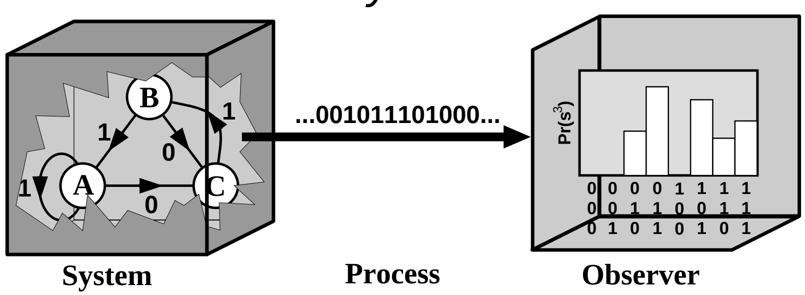
- Our estimates for h_μ and $\mathbf{E}$ depend strongly on x_c.
- Using an $x_c \neq 0.5$ leads to an h_μ is always lower than the true value.
- Using an $x_c \neq 0.5$ can lead to an over- or an under-estimate of $\mathbf{E}$.



- Note: $r = 3.8$ in this figure.

Example III: A Randomness Puzzle

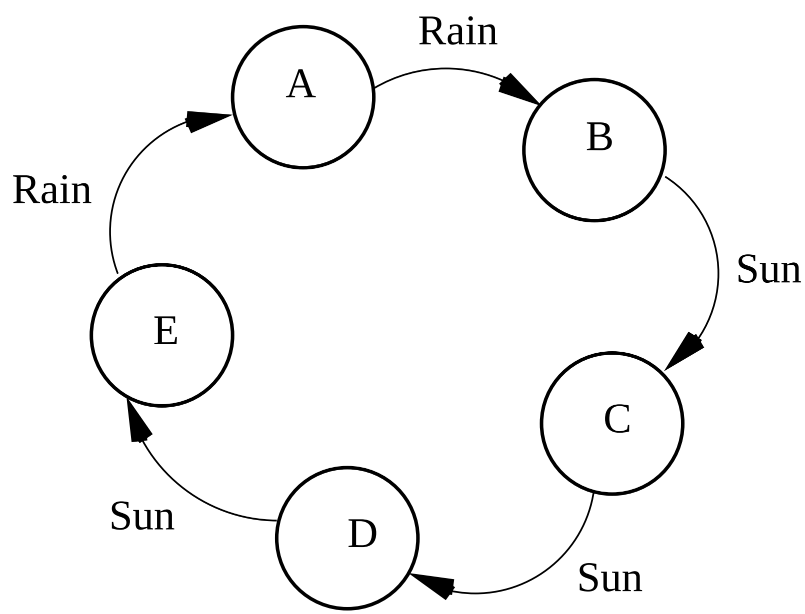
- Suppose we consider the binary expansion of π. Calculate its entropy rate h_μ and we'll find that it's 1.
- How can π be random? Isn't there a simple, deterministic algorithm to calculate digits of π?
- It is not random if one uses Kolmogorov complexity, since there is a short algorithm to produce the digits of π.
- It is random if one uses histograms and builds up probabilities over sequences.
- This points out the *model-sensitivity* of both randomness and complexity.



- Histograms are a type of model. See, e.g., Knuth. arxiv.org/physics/0605197. 2006.

Example IV: Unpredictability due to Asynchrony

- Imagine a strange island where the weather repeats itself every 5 days. It's rainy for two days, then sunny for three days.



- You arrive on this deserted island, ready to begin your vacation. But, you don't know what day it is: $\{A, B, C, D, E\}$.
- Eventually, however, you will figure it out.

Example IV: Unpredictability due to Asynchrony

- Once you are synchronized—you know what day it is—the process is perfectly predictable; $h_\mu = 0$.
- However, before you are synchronized, you are uncertain about the internal state. This uncertainty decreases, until reaching zero at synchronization.
- Denote by $\mathcal{H}(L)$ the average state uncertainty after L observations are made.
- The total state uncertainty experienced while synchronizing is the **Transient Information $\mathbf{T}$**:

$$\mathbf{T} \equiv \sum_{L=0}^{\infty} \mathcal{H}(L) \ . \tag{8}$$

Example IV: Unpredictability due to Asynchrony

- It turns out that different periodic sequences with the same P can have very different $\mathbf{T}$'s.
- For a given period P:

$$\mathbf{T}_{\max} \sim \frac{P}{2} \log_2 P \;, \tag{9}$$

and

$$\mathbf{T}_{\min} \sim \frac{1}{2} \log_2^2 P \;, \tag{10}$$

- E.g., if $P = 256$, then

$$\mathbf{T}_{\max} \approx 1024 \;, \text{ and } \; \mathbf{T}_{\min} \approx 32 \;. \tag{11}$$

- For disturbingly more detail, see Feldman and Crutchfield, “Synchronizing to Periodicity.” *Advances in Complex Systems.* 7:329-355. 2004.

Summary of Examples

- In all cases choice of representation and the state of knowledge of the observer influence the measurement of entropy or complexity.
  1. Ignored complexity is converted to entropy.
  2. Measurement choice can lead to an underestimate of h_μ and an over- or under-estimate of $\mathbf{E}$.
  3. π appears random.
  4. A periodic sequence is unpredictable and, in a sense, complex.
- Hence, statements about unpredictability or complexity are necessarily a statement about the observer, the observed, and the relationship between the two.
- So complexity and entropy are relative, but in an objective, clearly specified way.

Modeling Modeling

- Much of what I have presented in the last several lectures can be viewed as an abstraction of the modeling process itself.
- These examples provide a crisp setting in which one can explore trade-offs between, say, the complexity of a model and the observed unpredictability of the object under study.
- The choice of model can strongly influence the result yielded by the model. This influence can be understood.
- The hope is these models of modeling can give us some general, qualitative insight into modeling.

Model Dependence

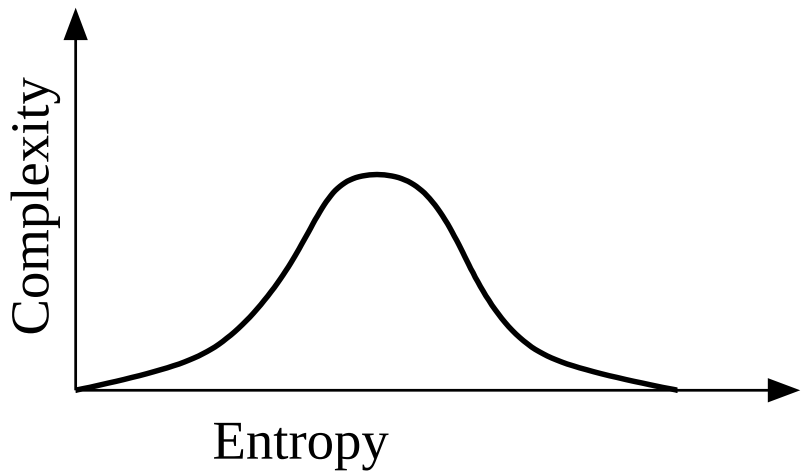
- There is no (computable), all-purpose measure of randomness or complexity.
- This isn't cause for despair. Just be as clear as you can about your modeling assumptions.
- Sometimes modeling assumptions can be hidden.
- I don't think will ever be a 100% objective measure of complexity. A statement about complexity will always be, to some extent, a statement about both the observer and the observed.

Complexity vs. Entropy

- What is the relationship between complexity and entropy?
- Are they completely unrelated? Is complexity the opposite of entropy?
- Is complexity an *absence* of unpredictability, or the *presence* of something else?

One approach: Prescribing Complexity vs. Entropy Behavior

- Zero Entropy $\longrightarrow$ Predictable $\longrightarrow$ simple and not complex.
- Maximum Entropy $\longrightarrow$ Perfectly Unpredictable $\longrightarrow$ simple and not complex.
- Complex phenomena combine order and disorder.
- Thus, it must be that complexity is related to entropy as shown:

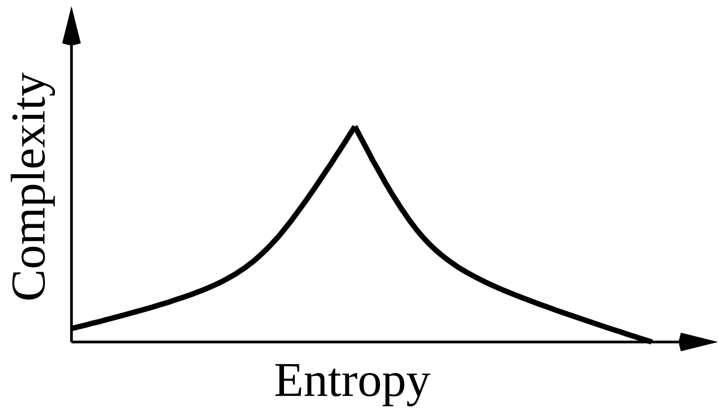


- This plot is often used as the central criteria for defining complexity.

Complexity-Entropy Phase Transition?

Edge of Chaos?

- Additionally, it has been conjectured that there is a sharp transition in complexity as a function of entropy:

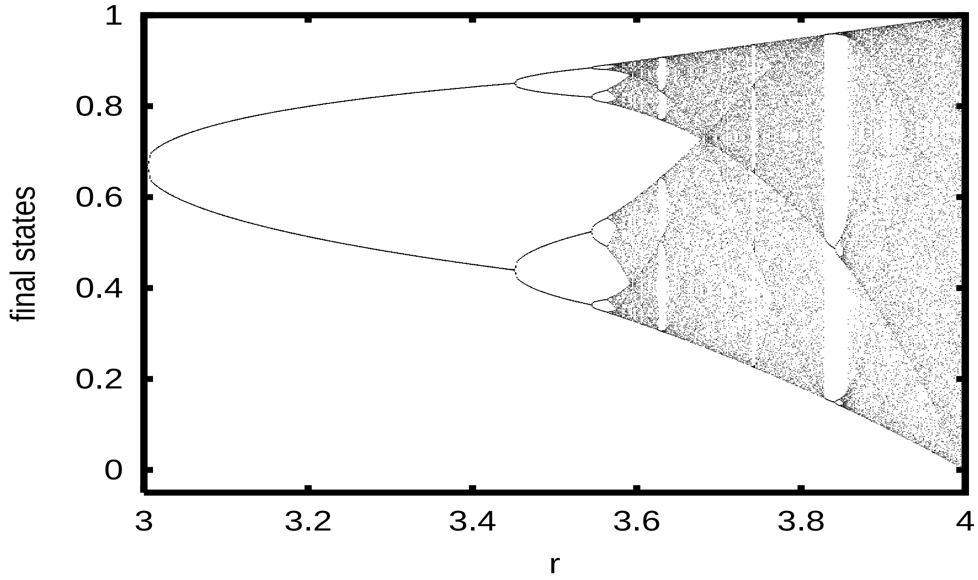


- Perhaps this complexity-entropy curve is *universal*—it is the same for a broad class of apparently different systems.
- Part of the motivation for this is the remarkable success of universality in critical phenomena and condensed matter physics.

Complexity vs. Entropy: A Different Approach
Define Complexity on its own Terms

- Do not prescribe a particular complexity-entropy behavior.
- To be useful, a complexity measure must have a clear interpretation that accounts in a direct way for the correlations and organization in a system.
- Consider a well known complexity measures: excess entropy
- Calculate complexity and entropy for a range of model systems.
- Plot complexity vs. entropy. This will directly reveal how complexity is related to entropy.
- Is there a universal complexity-entropy curve?

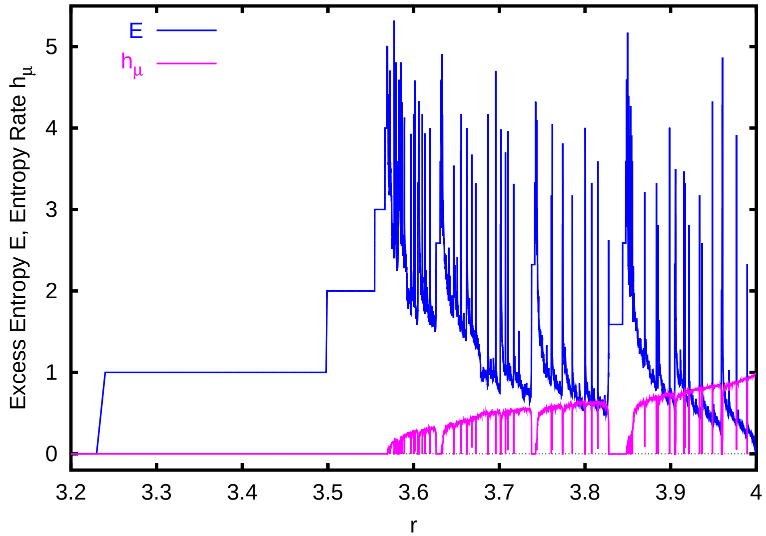
Logistic Equation: Bifurcation Diagram



- For a given r (horizontal axis), the “final states” are shown.
- Chaotic behavior appears as a solid vertical line.
- Examples:
  - $r = 3.2$: Period 2.
  - $r = 3.5$: Period 5.
  - $r = 3.7$: Chaotic.

Complexity vs. Entropy: Logistic Equation

Plot the excess entropy $\mathbf{E}$ and the entropy rate h_μ for the logistic equation as a function of the parameter r.



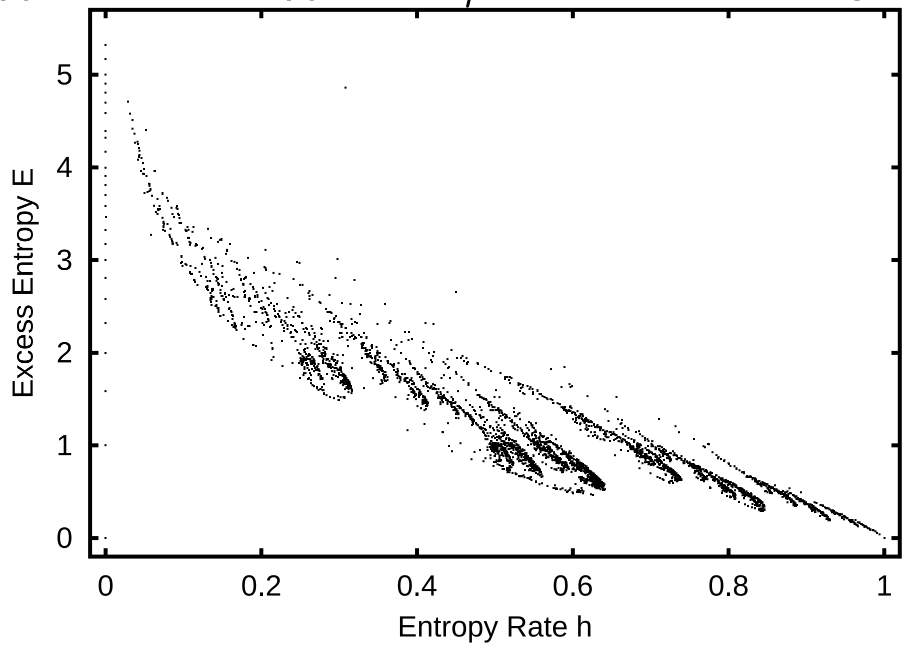
- Note that $\mathbf{E}$ and h_μ depend on a complicated way on r.
- Hard to see how complexity and entropy are related.
- Numerical results. For each r, 1×10^7 symbols were generated. The largest L was 30 for low entropy sequences. r was varied by increments of 0.0001.

Complexity-Entropy Diagrams

- Plot complexity vs. entropy. This will directly reveal how complexity is related to entropy.
- This is similar to the idea behind phase portraits in differential equations: plot two variables against each other instead of as a function of time. This shows how the two variables are related.
- It provides a parameter-free way to look at the intrinsic information processing of a system.
- Complexity-entropy plots allow comparisons across a broad class of systems.

Complexity-Entropy Diagram for Logistic Equation

- Excess entropy $\mathbf{E}$ vs. entropy rate h_μ from two slides ago.



- Structure is apparent in this plot that isn't visible in the previous one.
- Not all complexity-entropy values can occur; there is a forbidden region.
- Maximum complexity occurs at zero entropy.
- Note the self-similar structure. This isn't surprising, since the bifurcation diagram is self-similar.

Ising Models

Consider a one- or two-dimensional Ising system with nearest and next nearest neighbor interactions:

- This system is a one- or two-dimensional lattice of variables $s_i \in \{\pm 1\}$.
- The energy of a configuration is given by:

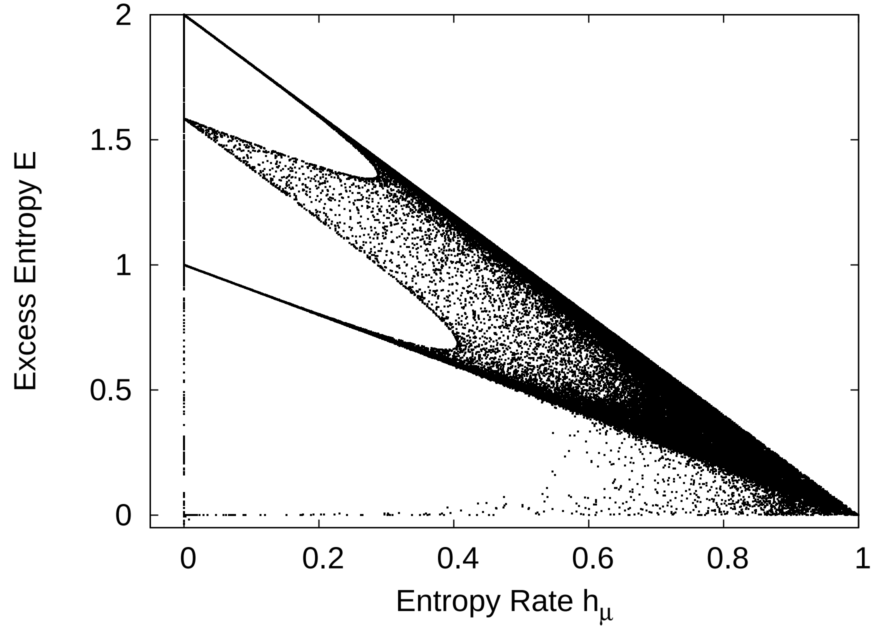
$$\mathcal{H} \equiv -J_1 \sum_i s_i s_{i+1} - J_2 \sum_i s_i s_{i+2} - B \sum s_i \ .$$

- The probability of observing a configuration $\mathcal{C}$ is given by the Boltzmann distribution:

$$\Pr(\mathcal{C}) \propto e^{-\frac{1}{T}\mathcal{H}(\mathcal{C})} \ .$$

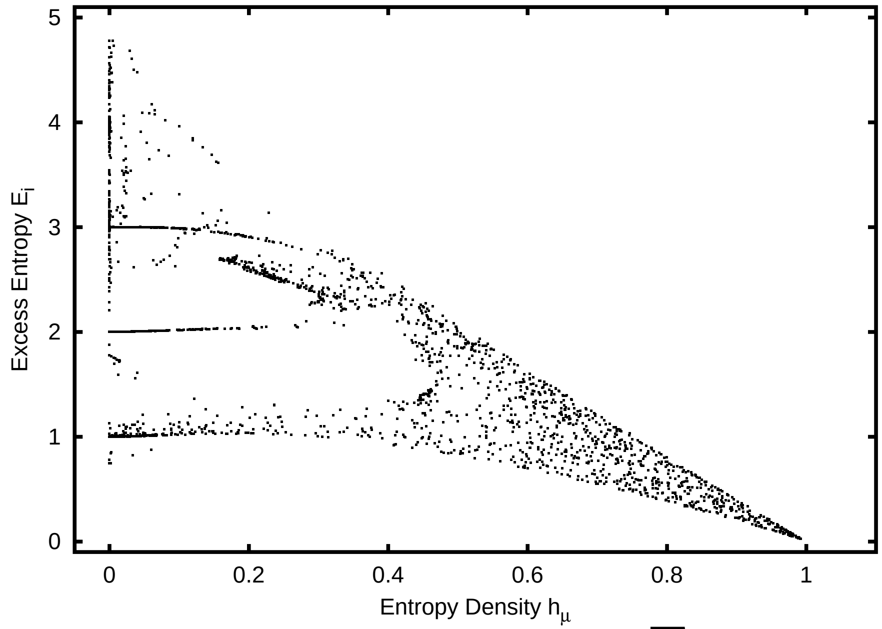
- Ising models are very generic models of spatially extended, discrete degrees of freedom that have some interaction that makes them want to either do the same or the opposite thing.

Complexity-Entropy Diagram for 1D Ising Models



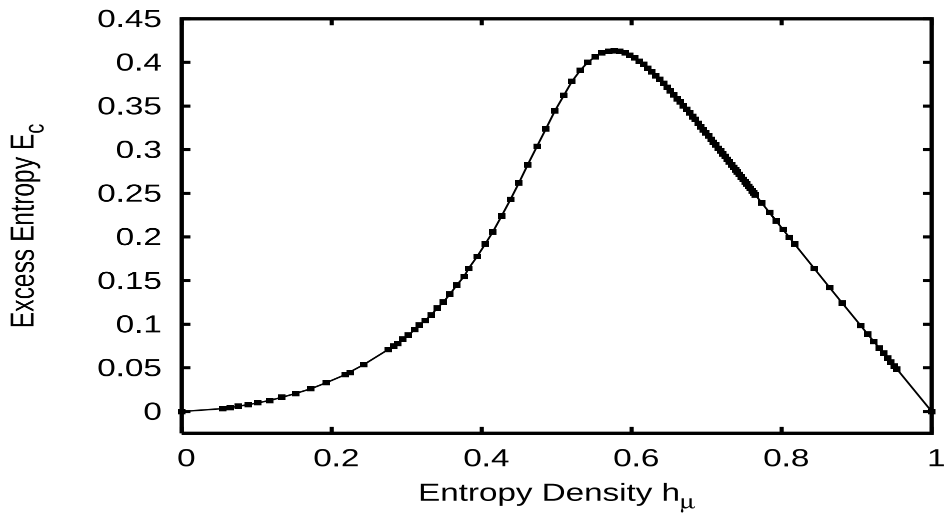
- Excess entropy $\mathbf{E}$ vs. entropy rate h_μ for the one-dimensional Ising model with anti-ferromagnetic couplings.
- Model parameters are chosen uniformly from the following ranges: $J_1 \in [-8, 0]$, $J_2 \in [-8, 0]$, $T \in [0.05, 6.05]$, and $B \in [0, 3]$.
- Note how different this is from the logistic equation.
- These are exact transfer-matrix results.

Complexity-Entropy Diagram for 2D Ising Models



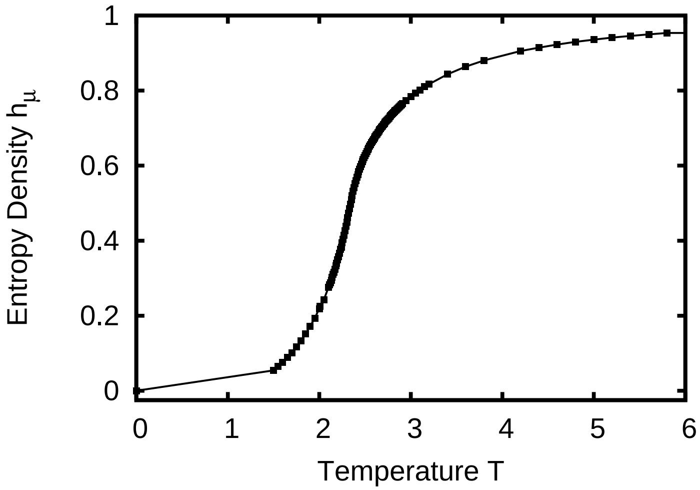
- Mutual information form of the excess entropy $\mathbf{E}_\mathrm{i}$ vs. entropy density h_μ for the two-dimensional Ising model with AFM couplings
- Model parameters are chosen uniformly from the following ranges: $J_1 \in [-3, 0]$, $J_2 \in [-3, 0]$, $T \in [0.05, 4.05]$, and $B = 0$.
- Surprisingly similar to the one-dimensional Ising model.
- Results via Monte Carlo simulation of 100x100 lattices.

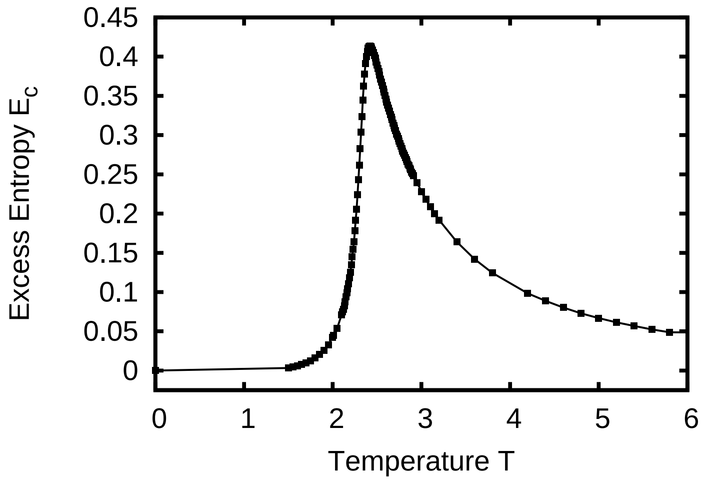
Complexity-Entropy Diagram for 2D Ising Model Phase Transition



- Convergence form of the excess entropy $\mathbf{E}_c$ vs. entropy density h_μ for the two-dimensional Ising model with NN couplings and no external field.
- Model undergoes phase transition as T is varied at $T \approx 2.269$.
- There is a peak in the excess entropy, but it is somewhat broad.
- Results via Monte Carlo simulation of 100x100 lattice.

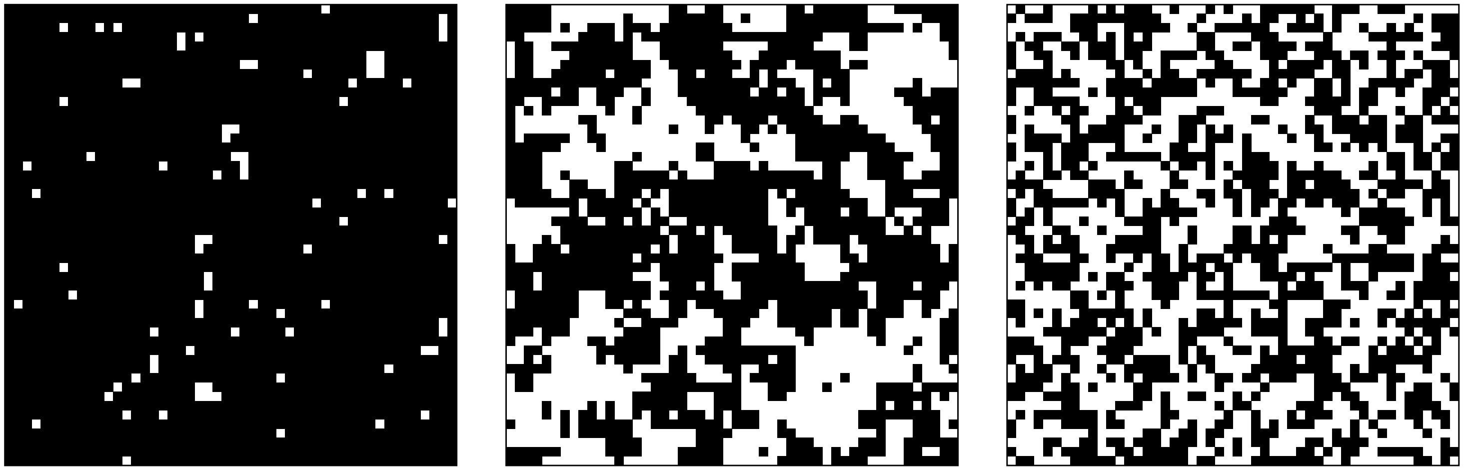
Complexity-Entropy Diagram for 2D Ising Model Phase Transition, continued





- Convergence form of the excess entropy $\mathbf{E}_c$ vs. entropy density h_μ versus temperature T for the two-dimensional Ising model with NN couplings and no external field.
- Model undergoes phase transition as T is varied at $T \approx 2.269$.
- There is a peak in the excess entropy is broader if plotted as a function of T than when plotted against h_μ as on the previous slide.
- Results via Monte Carlo simulation of 100x100 lattice.

Ising Model Configurations

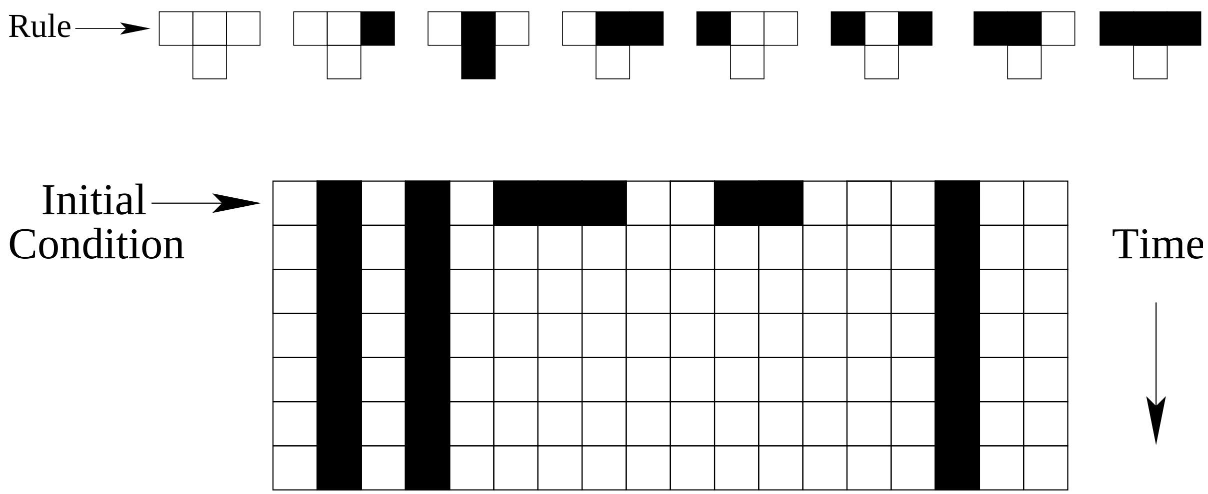


- Typical configurations for the 2D Ising model below, at, and above the critical temperature.

Cellular Automata

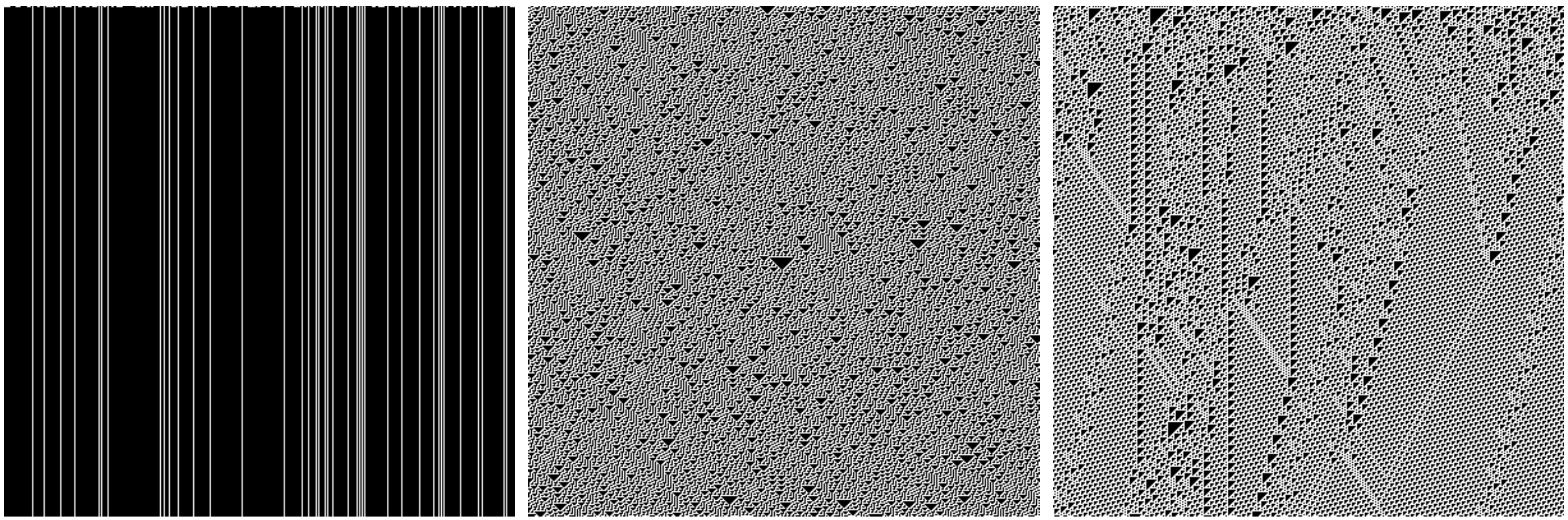
- The next row in the grid is determined by the row directly above it according to a given rule
- Start with a random initial condition

Example:



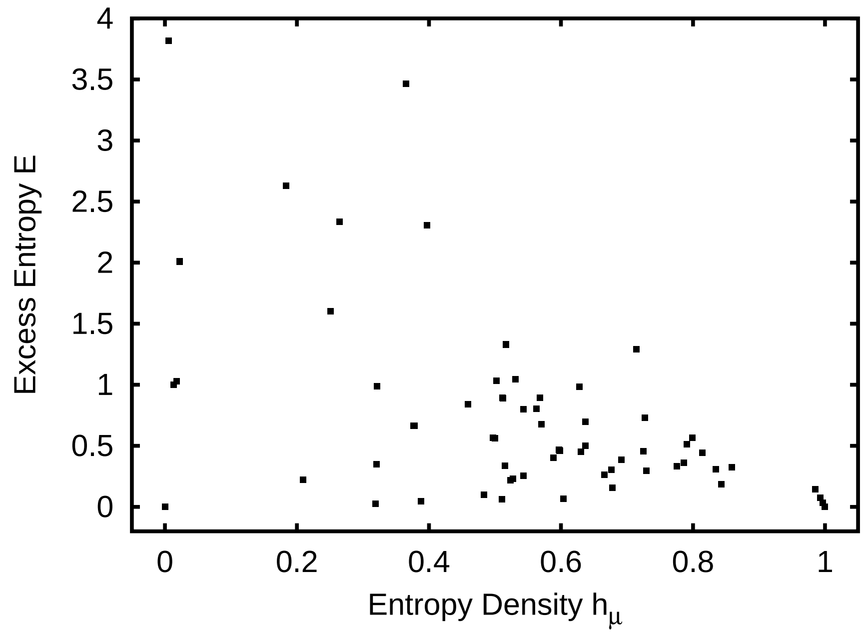
- The number of cells away from the center cell that the rule considers is known as the radius of the CA.

Different Rules Yield Different Patterns



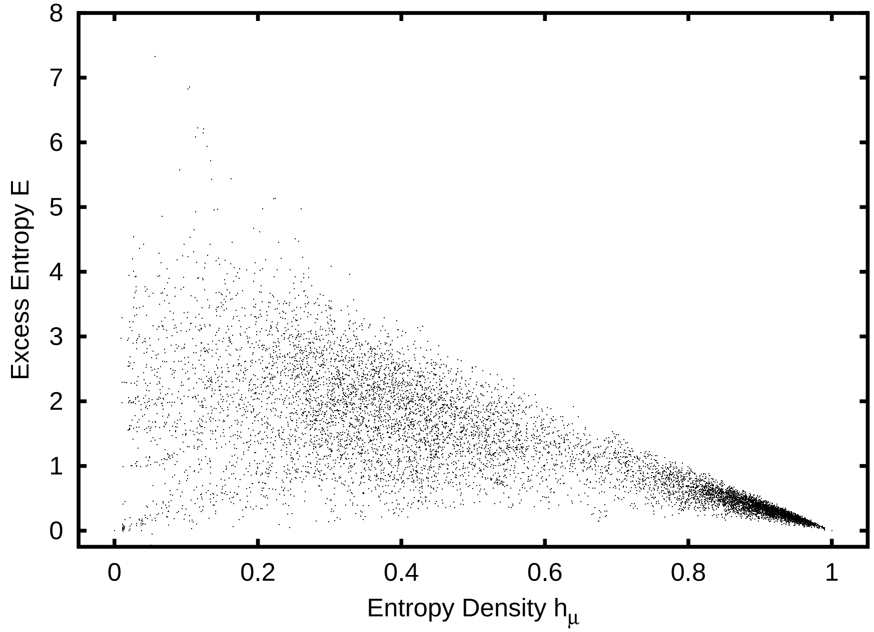
- Each pattern is for a different rule.

Complexity-Entropy Diagram for Radius-1, 1D CAs (aka Elementary CAs, or ECAs)



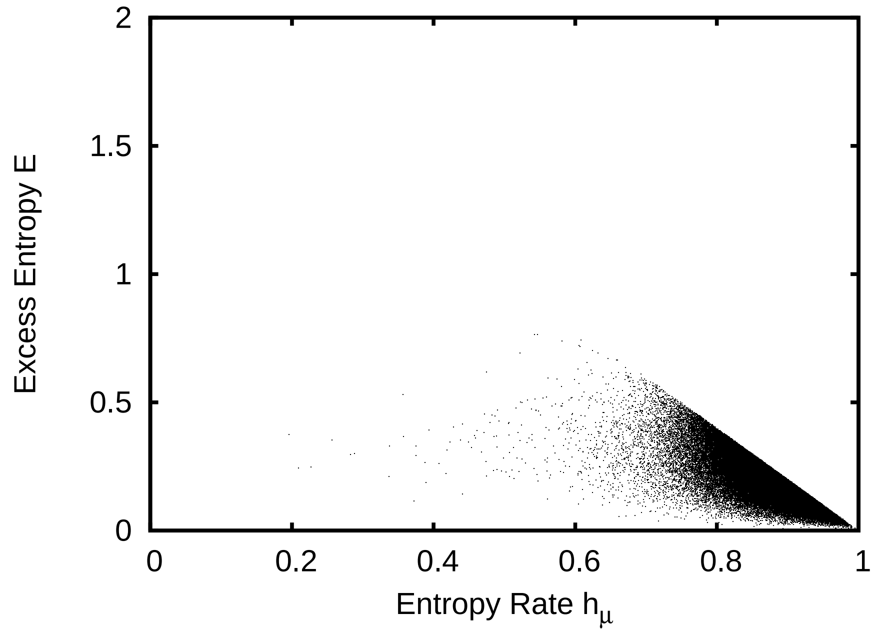
- Excess entropy $\mathbf{E}$ and entropy density h_μ for all distinct (88) one-dimensional elementary cellular automata.
- $\mathbf{E}$ and h_μ from the spatial strings produced by the CAs.
- Since there are so few ECAs, it's hard to discern a pattern. What if we try radius-2 CAs?

Complexity-Entropy Diagram for Radius-2, 1D CAs



- Excess entropy $\mathbf{E}$ vs. entropy rate h_μ for $10,000$ radius-2, binary CAs.
- $\mathbf{E}$ and h_μ from the spatial strings produced by the CAs.
- The CAs were chosen uniformly from the space of all such CAs.
- There are around 4.3×10^9 such CAs, so it is impossible to sample the entire space.

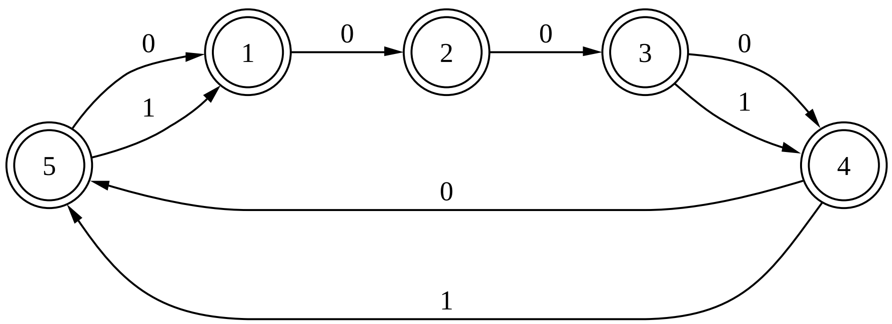
Complexity-Entropy Diagram for Markov Models

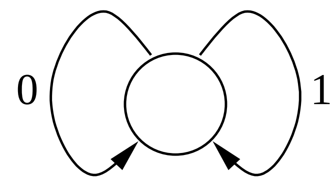


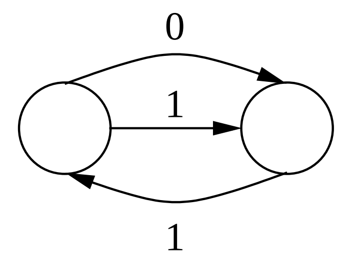
- Excess entropy $\mathbf{E}$ vs. entropy rate h_μ for $100,000$ random Markov models.
- The Markov models here have four states, corresponding to dependence on the previous two symbols, as in the 1D NNN Ising model.
- Transition probabilities chosen uniformly on $[0,1]$ and then normalized.
- Note that these systems have no forbidden sequences.

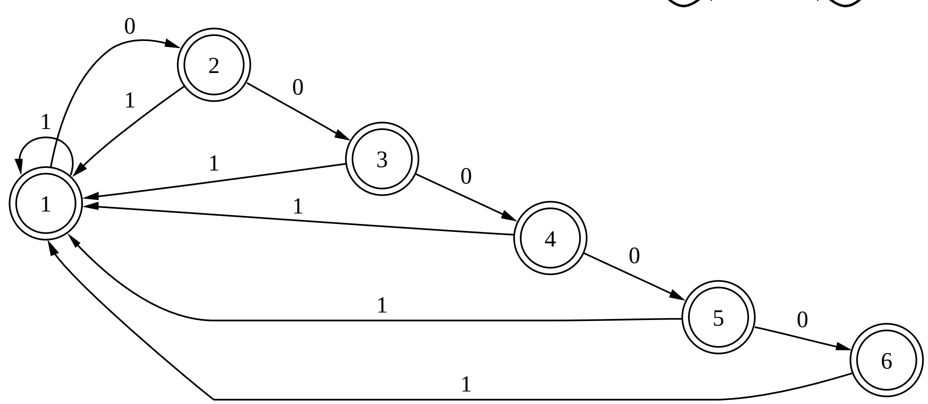
Topological Markov Chain Processes

- Consider finite-state machines that produce 0's and 1's.
- Assume all branching transitions are equally probable
- Examples:





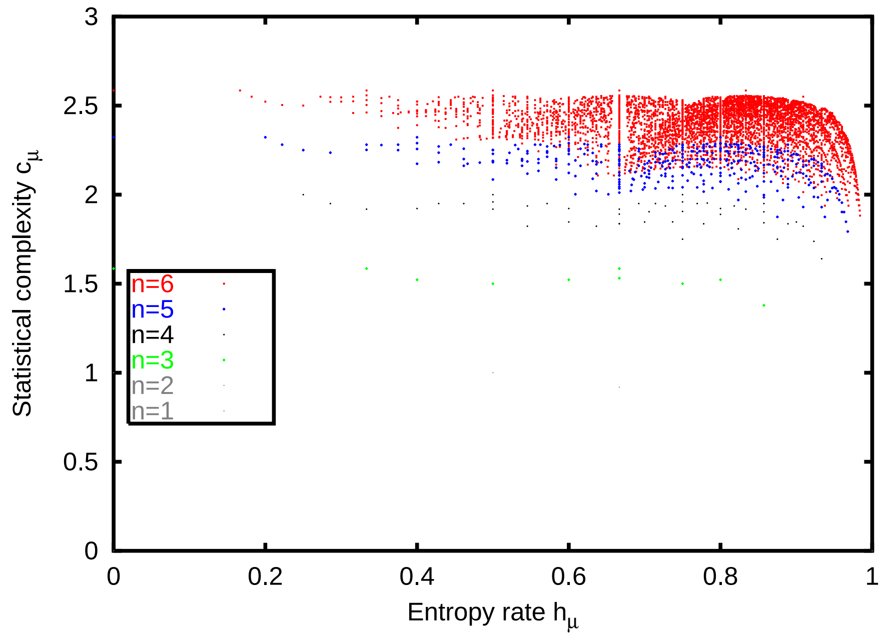




Topological Processes and Statistical Complexity

- These topological processes can be exhaustively enumerated for any finite number of states.
- We now use a different measure of complexity: the *statistical complexity* C_μ
- C_μ is the Shannon entropy of the asymptotic distribution over states.
- We consider only minimal machines.
- $C_\mu \geq \mathbf{E}$.

Complexity-Entropy Diagram for Topological Processes



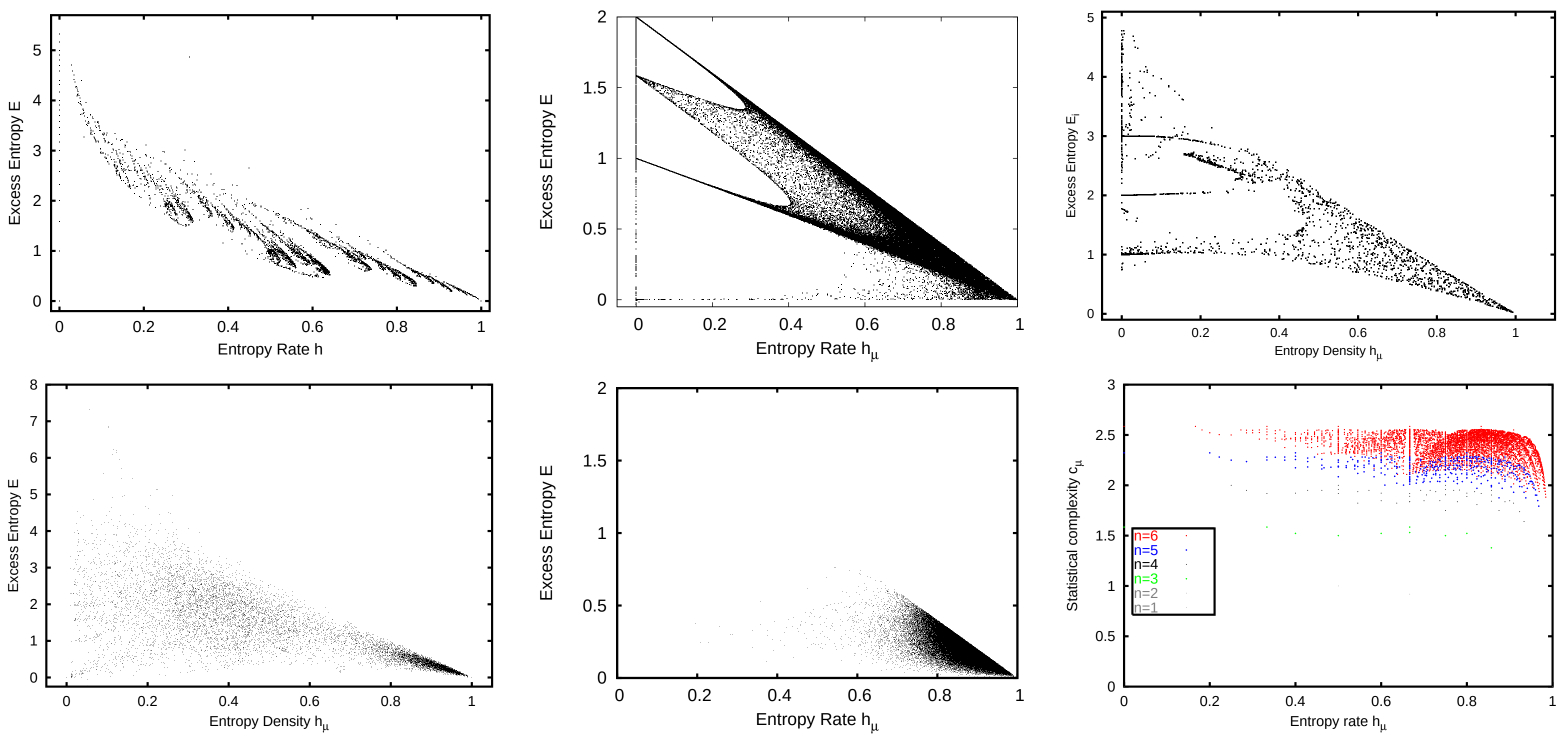
- h_μ, C_μ pairs for all $14,694$ distinct topological processes of $n = 1$ to $n = 6$ states. (Work done by Carl McTague.)
- Note the prevalence of high-entropy, high-complexity processes.

A Gallery of Complexity-Entropy Diagrams

The next slide shows, left to right, top to bottom, complexity-entropy diagrams for:

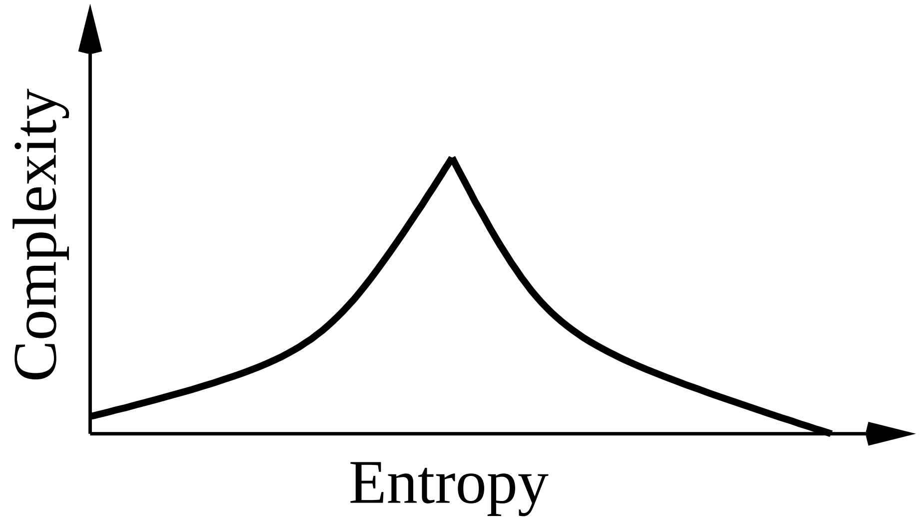
1. Logistic Equation
2. One-Dimensional Ising model with nearest- and next-nearest-neighbor interactions
3. Two-Dimensional Ising model with nearest- and next-nearest-neighbor interactions
4. One-Dimensional radius-2 cellular automata
5. Random Markov chains
6. All 6-state topological processes

A Mosaic of Complexity-Entropy Diagrams



Complexity-Entropy Diagrams: Summary

- Is it the case that there is a universal complexity-entropy diagram?



- No!
- However, because of this non-universality, complexity-entropy diagrams provide a useful way to compare the information processing abilities of different systems.
- Complexity-entropy plots allow comparisons across a broad class of systems.

Complexity-Entropy Diagrams: Conclusions

- There is not a universal complexity-entropy curve.
- Complexity is not necessarily maximized at intermediate entropy values.
- It is not always the case that there is a sharp complexity-entropy transition.
- Complexity-entropy diagrams provide a way of comparing the information processing abilities of different systems in a parameter-free way.
- Complexity-entropy diagrams allow one to compare the information processing abilities of very different model classes on similar terms.
- There is a considerable diversity of complexity-entropy behaviors.

Some Thoughts on the Past, Present, and Future of Complexity Measures

- Over the past two decades there have been considerable advances in how we think about and measure complexity, memory, structure, and pattern.
- There are now several, well understood and (fairly) widely used ways to approach structural complexity. Useful for:
  - Analyzing real data
  - Deepening understanding of model systems and fundamental sources of complexity or regularity.
  - Shedding light on foundational issues in pattern discovery.
- Along the way there has been (too much) hype and quite a few neat ideas that have turned out to be not as useful as one may have hoped.

A Few Cautionary Notes

- The term *complexity* has many different meanings. At least one adjective is needed to help distinguish between different uses of the word.
- Be cautious of "edge of chaos" hype.
- Don't invent a new complexity measure unless you have a compelling reason to do so.
- A good complexity measure should tell you something other than the value of the complexity measure.
- All Universal-Turing-Machine-based complexity measures suffer from several drawbacks:
  1. They are uncomputable.
  2. By adopting a UTM, the most powerful discrete computation model, one loses the ability to distinguish between systems that can be described by computational models less powerful than a UTM.

Complexity = Order $\times$ Disorder?

- There are a number of complexity measures of the form:

$$\text{Complexity} = \text{Order} \times \text{Disorder}$$

- Disorder is usually some form of entropy.
- Sometimes "order" is simply $(1 - h_\mu)$.
- Often, "order" is taken to be some measure of "distance from equilibrium," where equilibrium and equiprobability are sometimes considered to be synonymous.

In my view these sorts of complexity measures have some serious shortcomings:

- Lack a clear interpretation and direct accounting of structure.
- Unclear that distance from equilibrium is equivalent to order.
- Assign a value of zero complexity to all systems with vanishing entropy.

Open Questions and Future Directions

1. Mathematical and Conceptual Foundations.
   (a) Ay's and Löhr's talks today
   (b) Situations in which the excess entropy and/or the statistical complexity diverge
2. Extensions
   (a) Non-stationary data
   (b) Two-dimensional systems
      - Feldman and Crutchfield, *Physical Review E*,**67**:051104, 2003 and references therein.
      - Shalizi, et al., *Phys. Rev. Lett.* **93**:118701, 2004.
      - Young, et al. *Physical Review Letters.* **94**:098701. 2005.
      - Shalizi, et al. *Phys. Rev. E.* **73**: 036104. 2006.
   (c) Complexity of networks

Open Questions and Future Directions

3. Applications

   (a) Understand more fully the relation between various complexity measures and critical phenomena.

   (b) Disordered or inhomogeneous systems, e.g. spin glasses.

   (c) Agent-based models.

   (d) Empirical data, a.k.a., the real world. (Watkins' talk.)

   (e) Other model systems. (Nerukh's talk.)

4. Inference

   (a) Better estimators for causal states, statistical complexity, etc.

   (b) Connection between measures of complexity and the difficulty of learning a pattern.

   (c) On-line complexity estimation.

Open Questions and Future Directions

- In general, I believe that these tools are a useful framework for considering questions of complexity, organization, and emergence.
- These concerns seem to me to be central to the study of complex systems.

Thanks and Acknowledgments

- Much of what I've presented is joint work with with Jim Crutchfield.
- Thanks also to: Hao Bai-lin, Erica Jen, Kristian Lindgren, Susan McKay, Carl McTague, Cris Moore, Richard Scalettar, Cosma Shalizi, Dan Upper, Dowman Varn, Jon Wilkins, Karl Young,

- Graduate Students: Please consider applying to the Santa Fe Institute's Complex Systems Summer Schools in Beijing, China, and Santa Fe, USA.
- Please also consider applying for SFI Postdoctoral Fellow positions.
- I would welcome comments, questions, suggestions, and critique.
- Thank you!