Information Theory: Part II
Applications to Stochastic Processes

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

$$\cdots 0011001001010110100110011101 0110 \cdots$$

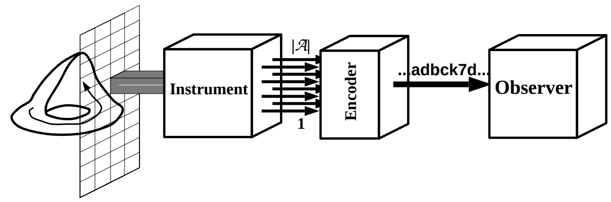
- In so doing, we will be led to two important quantities
  1. **Entropy Rate:** The irreducible randomness of the system.
  2. **Excess Entropy:** A measure of the complexity of the sequence.

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"): $\overrightarrow{S} \equiv S_0\, S_1\, S_2 \cdots$

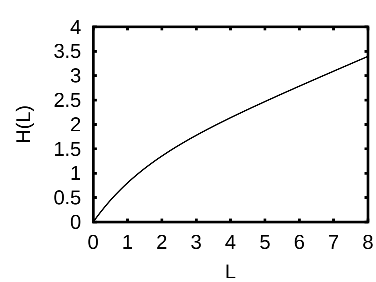
$$\cdots 11010100101101010101001001010010 \cdots$$

Entropy Growth

- Entropy of L-block:

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

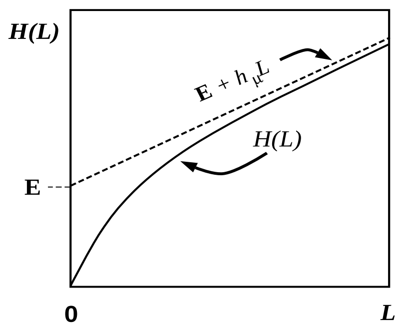
- $H(L) =$ average uncertainty about the outcome of L consecutive variables.



- $H(L)$ 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:*

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

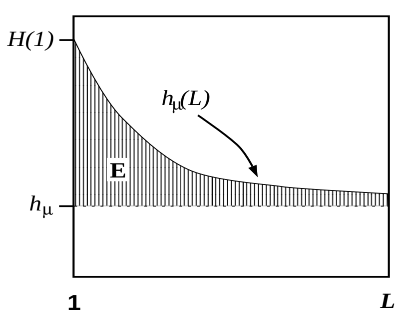
- $h_\mu(L) = H[S_L | S_1 S_1 \ldots S_{L-1}]$
- I.e., $h_\mu(L)$ is the average uncertainty of the next symbol, given that the previous L symbols have been observed.

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_μ = 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** 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":

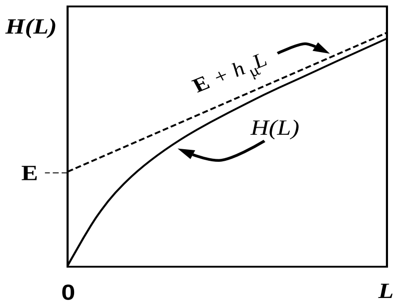
$$\mathbf{E} = I(\overleftarrow{S};\overrightarrow{S}) \equiv \sum_{\{\overleftrightarrow{s}\}} \Pr(\overleftrightarrow{s}) \log_2 \left[\frac{\Pr(\overleftrightarrow{s})}{\Pr(\overleftarrow{s})\Pr(\overrightarrow{s})} \right] .$$

- $\mathbf{E}$ 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

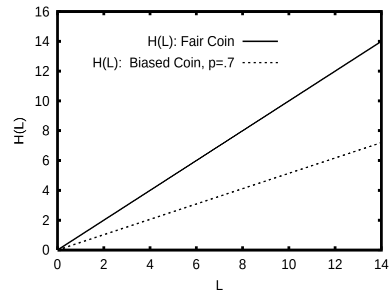
- $\mathbf{E}$ is the y-intercept of the straight line to which $H(L)$ asymptotes.
- $\mathbf{E} = \lim_{L\to\infty} [H(L) - h_\mu L]$.



Excess entropy summary:

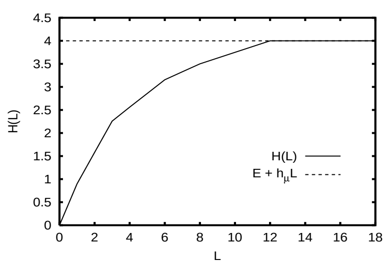
- Is a structural property of the system — measures a feature complementary to entropy.
- Measures memory or spatial structure.
- Lower bound for statistical complexity, minimum amount of information needed for minimal stochastic model of system

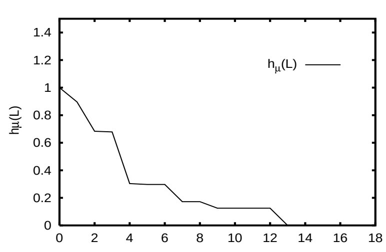
Example I: Fair Coin



- 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.

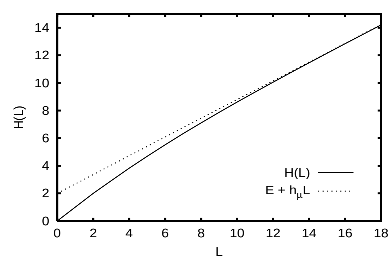
Example II: Periodic Sequence

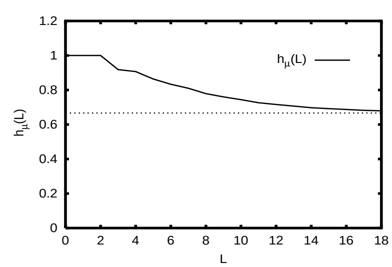




- Sequence: $\ldots 1010111011101110 \ldots$
- $h_\mu \approx 0$; the sequence is perfectly predictable.
- $\mathbf{E} = \log_2 16 = 4$: four bits of phase information
- For any period-p sequence, $h_\mu = 0$ and $\mathbf{E} = \log_2 p$.

Example III: Random, Random, XOR





- Sequence: two random symbols, followed by the XOR of those symbols.
- $h_\mu = \frac{2}{3}$; two-thirds of the symbols are unpredictable.
- $\mathbf{E} = \log_2 4 = 2$: two bits of phase information.
- For many more examples, see Crutchfield and Feldman, Chaos, 15: 25-54, 2003.

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

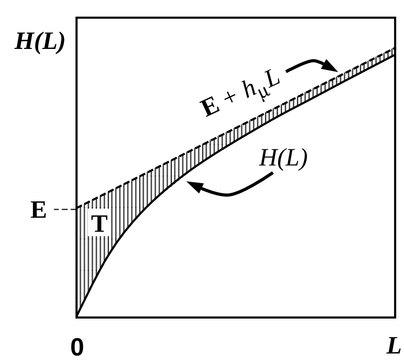
- Crutchfield and Packard, *Intl. J. Theo. Phys*, 21:433-466. (1982); *Physica D*, 7:201-223, 1983. [Dynamical systems]
- Shaw, "The Dripping Faucet ..., " Aerial Press, 1984. [A dripping faucet]
- Grassberger, *Intl. J. Theo. Phys*, 25:907-938, 1986. [Cellular automata, dynamical systems]
- Szépfalusy and Györgyi, *Phys. Rev. A*, 33:2852-2855, 1986. [Dynamical systems]
- Lindgren and Nordahl, *Complex Systems*, 2:409-440. (1988). [Cellular automata, 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]
- Feldman and Crutchfield. *Physical Review E*, 67:051104. 2003. [Two-dimensional Ising models]

Excess Entropy: Selected References and Applications, continued

- 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]

Transient Information $\mathbf{T}$

- $\mathbf{T} \equiv \sum_{L=1}^{\infty} [\mathbf{E} + h_\mu L - H(L)]$.
- $\mathbf{T}$ is related to the total uncertainty experienced while synchronizing to a process.



- The shaded area is the transient information T.
- $\mathbf{T}$ measures how difficult it is to synchronize to a sequence.

Some Applications in Agent-Based Modeling Settings

1. If an agent doesn't have sufficient memory, its environment will appear more random. In a quantitative sense, regularities that are missed (as measured by the excess entropy) are converted into randomness (as measured by the entropy rate).

   Crutchfield and Feldman, Synchronizing to the Environment: Information Theoretic Constraints on Agent Learning. *Advances in Complex Systems.* 4. 251–264. 2001.

2. The average-case difficulty for an agent to synchronize to a periodic environment is measured by the transient information.

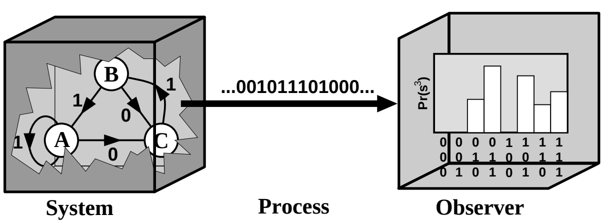
   Feldman and Crutchfield. Synchronizing to a Periodic Signal: The Transient Information and Synchronization Time of Periodic Sequences. *Advances in Complex Systems.* 7. 329–355. 2004.

Some Applications in Agent-Based Modeling Settings, continued

3. More generally it seems likely that the entropy and mutual information are useful tools for quantifying
   (a) properties of agents: e.g., how much memory they have
   (b) the behavior of agents: e.g, how unpredictably they act
   (c) properties of the environment: e.g., how structured it is

Estimating Probabilities

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



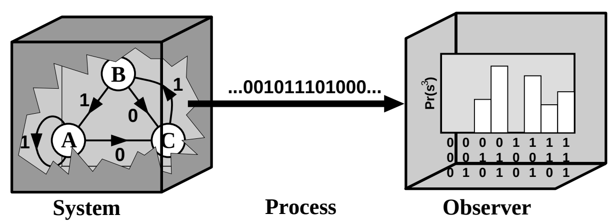
- One simply forms histograms of occurrences of particular sequences and uses these to estimate $\Pr(s^L)$, from which $\mathbf{E}$ and h_μ may be readily calculated.

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*?