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. In a sense, it is an opinion.
- 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

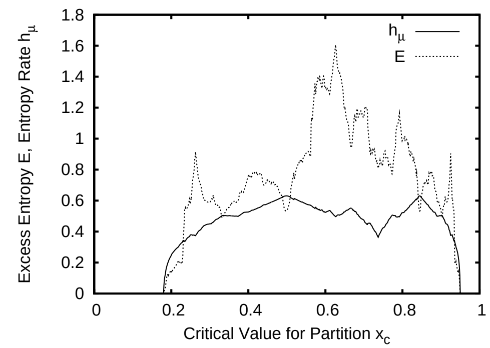
- Iterate the logistic equation: $x_{n+1} = f(x_n)$, where $f(x) = rx(1-x)$.
- Result is a sequence of numbers. E.g., $0.445, 0.894, 0.22, 0.344, \ldots$.
- Generate symbol sequence via:

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

- As we've seen, 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)

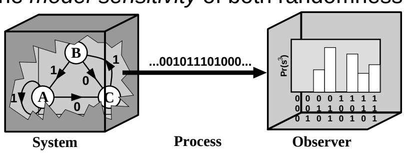
- Our estimates for h_μ and $\mathbf{E}$ depend strongly on x_c.
- Using an $x_c \neq 0.5$ leads to an h_μ is always lower than the true value.
- Using an $x_c \neq 0.5$ can lead to an over- or an under-estimate of $\mathbf{E}$.



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

Example III: A Randomness Puzzle

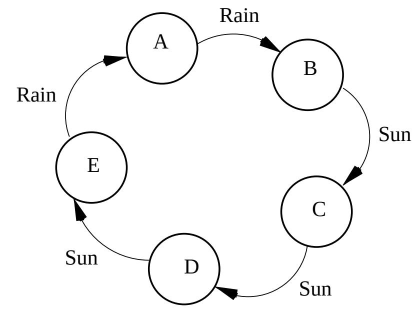
- Suppose we consider the binary expansion of π. Calculate its entropy rate h_μ and we'll find that it's 1.
- How can π be random? Isn't there a simple, deterministic algorithm to calculate digits of π?
- It is not random if one uses Kolmogorov complexity, since there is a short algorithm to produce 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** $\mathbf{T}$:

$$\mathbf{T} \equiv \sum_{L=0}^{\infty} \mathcal{H}(L) \ . \tag{1}$$

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:

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

and

$$\mathbf{T}_{\rm min} \sim \frac{1}{2} \log_2^2 P \ , \tag{3}$$

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

$$\mathbf{T}_{\rm max} \approx 1024 \ , \ \text{and} \ \ \mathbf{T}_{\rm min} \approx 32 \ . \tag{4}$$

- 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.