Lecture 8
Measuring Complexity
(with a side of information theory)

Logistics

- Current Reading
  - Gell-Mann, M. What is Complexity?, *Complexity, Vol 1, no. 1*, 1995 (Read by Feb 12th)
  - Crutchfield, J. Between Order and Chaos, *Nature Physics*, 2012 (Read by Feb 12th)
  - Mitchell. Chapter 3 and 4 by Feb 10^{th}.
  - Mitchell. Chapter 7 by Feb 17^{th}.

- Project 1/4 is due in 0 days at 6:00pm.

- Now you are a domain expert. Switch to reviewer mode.

A clarification

- Recall that Strange Attractors are chaotic.
- I have made the following two claims:

1. Strange Attractors cannot exist in systems with less than 3 dimensions.
2. the logistic map contains a strange attractor.

- What is the loophole that keeps me from being a liar?

A clarification

- Recall that Strange Attractors are chaotic.
- I have made the following two claims:

  1. Strange Attractors cannot exist in systems with less than 3 dimensions.
  2. the logistic map contains a strange attractor.

- Hint: we have seen two basic types of dynamical system. Statement 1 is true for one of these …

A clarification

- Recall that Strange Attractors are chaotic.
- I have made the following two claims:

1. Strange Attractors cannot exist in systems with less than 3 dimensions.
2. the logistic map contains a strange attractor.

- Hint: we have seen two basic types of dynamical system. Statement 1 is true for one of these …

Even more of a hint: Poincaré– Bendixson theorem (Lecture 2)

How should we measure the complexity of a system?

How to measure a system’s complexity?

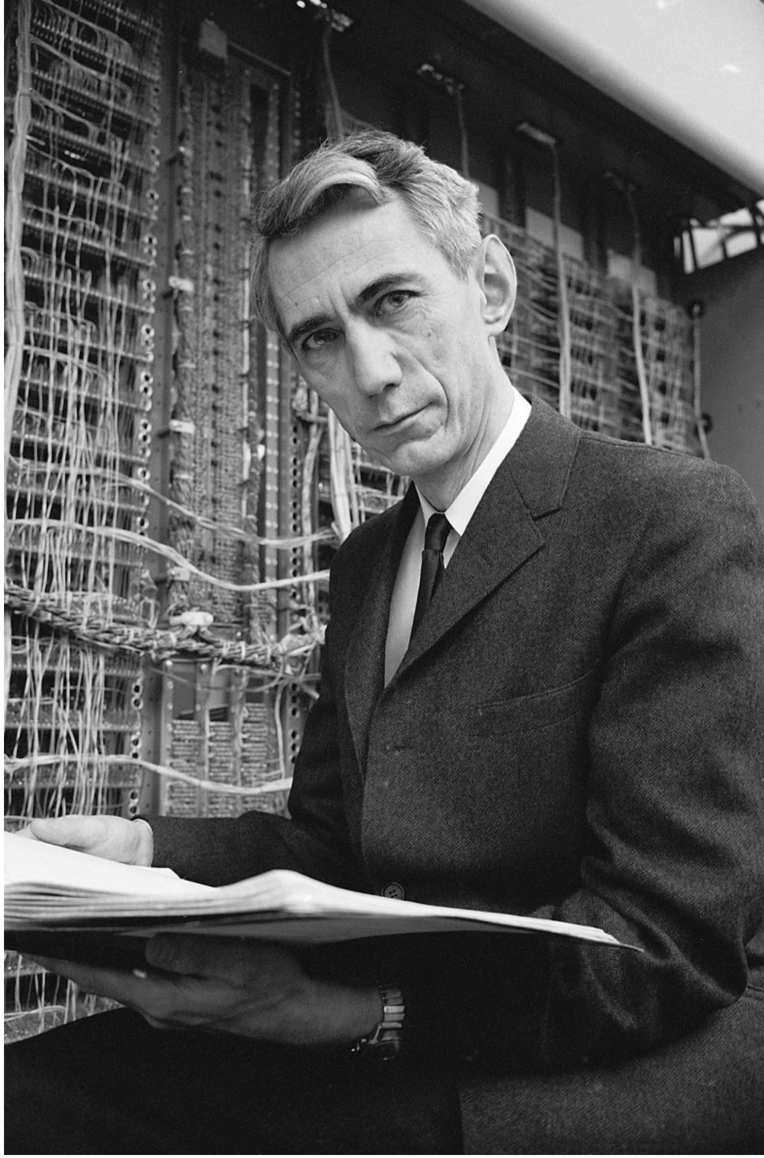
- Unpredictability, entropy
  - Information content.
- How difficult it is to describe
  - Length of the most concise description
  - No single model adequate to describe the system? (Lee Segel def. of complexity)
- Measuring how long until it halts, if ever
  - How long until it repeats itself?
- How difficult is it to construct?
  - Multiple levels of organization
  - Number of interdependencies

A Few Measures of Complexity (there are many)

- Asymptotic behavior of dynamical systems
  - Fixed points, limit cycles, chaos
  - Wolfram's CA classification, Langton's lambda parameter
- Computational complexity (Cook):
  - What resources does it take to compute a function?
- Language complexity
  - How complex a machine is needed to compute a function?
- Information-theoretic measures
  - Entropy, algorithmic complexity, mutual information
- Logical depth (Bennett)
  - Run-time of the smallest machine that generates the pattern and halts
- Thermodynamic depth (Lloyd and Pagels)
- Effective complexity (Gell-mann and Lloyd)

Information Theory

- Information
  - Decrease in uncertainty once we learn a certain value (symbol)
  - Additive: The more we learn the less uncertainty we have. Use logarithms to get this property
  - Expressed as bits
  - Normalized for probability of different symbols



Claude Shannon, Bell Labs

1948, "A Mathematical Theory of Communication"

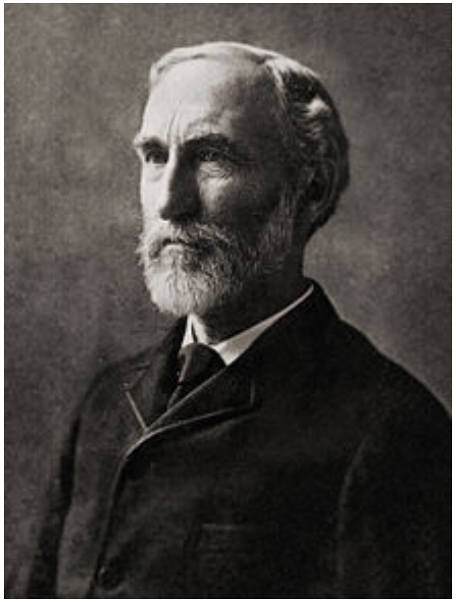
Cryptographer who worked with Alan Turing when Turing visited the US.

$$H = -\sum_i p_i \log_2(p_i)$$

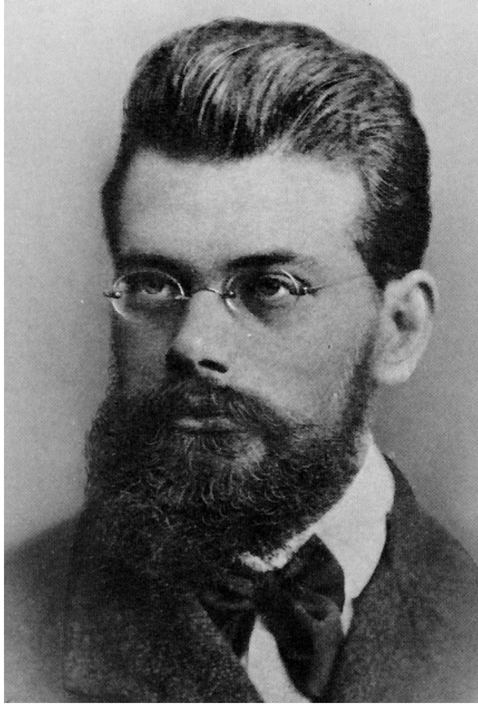
Information Theory

- Shannon´s information formula is identical to the formula for entropy in statistical thermodynamics.
- Entropy was defined by Bolzmann, Gibbs, and Planck in the late 19th C.
- This formula is the basis for the second law of thermodynamics (rules of energy)

$$S = -k_B \sum_i p_i \ln(p_i)$$



Willard Gibbs



Ludwig Boltzmann

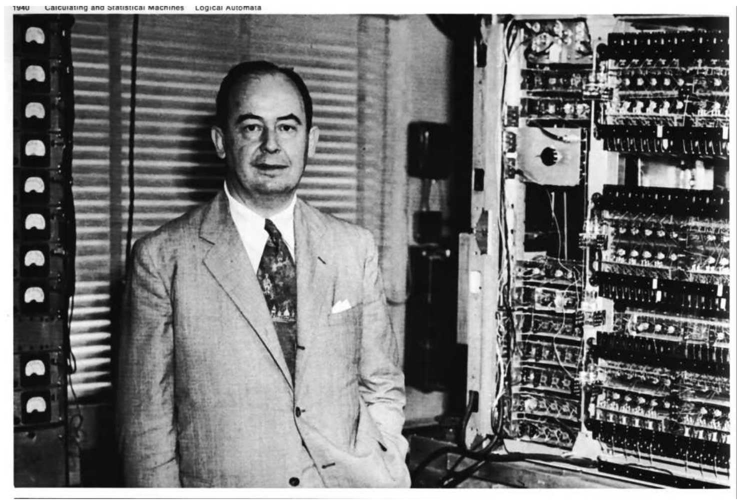
Max Planck

Entropy

- **You should call it entropy**, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. **In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage. – Jon von Neumann** *(Scientific American* Vol. 225 No. 3, (1971), p. 180.)

- The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work.

- **In mathematics you don't understand things. You just get used to them.**



Inventor of the Computer (along with Alan Turing) .

One of the lead scientists on the Manhattan project.

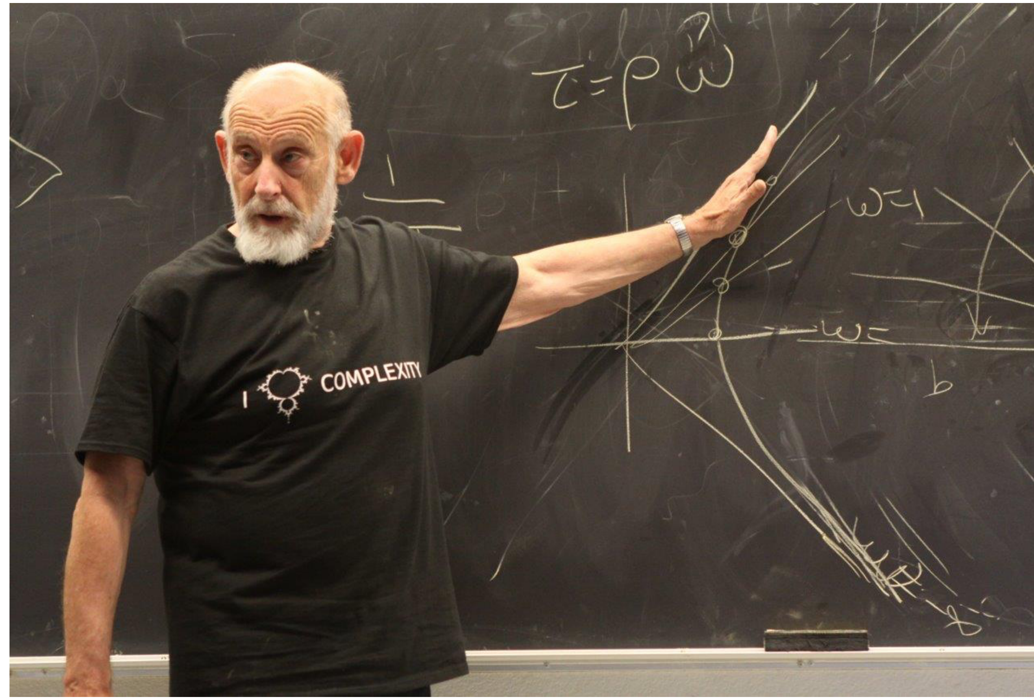
Inventor of Game Theory (which lead to strategy of *mutually assured destruction*).

Information content is a fundamental property of all systems.

The most likely configuration of a physical system is the one that Maximised the information entropy.

Entropy is a measure of how far the system is from being uniformly random.

In physical systems entropy Always increases leading to more randomness – less complexity.



Leonard Susskind, Stanford.
Known for his work in black hole entropy, string theory, and the holographic principle.

The holographic principle showed that information is not lost in a black hole
(contrary to Stephen Hawkins).

The most fundamental particle is the *bit.*

Susskind, Leonard (1995). "The World as a Hologram". *Journal of Mathematical Physics*. **36** (11): 6377–6396

Shannon Entropy

- Generalize previous equation to account for symbols appearing with different frequencies:

$$H = -\sum_i p_i \log_2(p_i)$$

- Entropy is measured in bits
- H measures the average uncertainty in the random variable
- Example: Consider a random variable with uniform distribution over 32 values
- Need 5-bit strings to label each outcome

Information and Entropy

- Information is defined to be the resolution of uncertainty.
- Resolution in the sense that the current uncertainty is removed by the receipt of information.
- It is the uncertainty before the information is obtained that defines the quantity of the information.

What is a random variable?

- A function defined on a sample space
  - Should be called *random function*
  - Independent variable is a point in the sample space, e.g., the outcome of an experiment.
- A function of outcomes, rather than a single outcome
- Probability distribution of the random variable X

- Example: $P\{X=x_j\} = f(x_j) \quad j = 1, 2, \ldots$
  - Toss 3 fair coins
  - Let X denote the number of heads that appear
  - X is a random variable taking on one of the values (0,1,2,3)
  - $P\{X=0\} = 1/8$; $p\{X=1\} = 3/8$; $p\{X=2\} = 3/8$; $p\{X=3\} = 1/8$

Slide courtesy of [Stephanie Forrest]

Example: Horse race

- Probabilities of 8 horses are:

$$\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64}\right)$$

- Calculate the entropy H of a message containing the outcome:

- Could send the index of the winning horse

  - How many bits of information would this take on average?

Slide courtesy of [Stephanie Forrest]

Example: Horse race

- Probabilities of 8 horses are:

$$\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64}\right)$$

- Calculate the entropy H of a message containing the outcome:

- Could send the index of the winning horse (3 bits)

- Or, use the following labels: 0, 10, 110, 1110, 111100, 111101, 111110, 111111
  - Avg. description length is 2 bits instead of 3

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4\frac{1}{64}\log\frac{1}{64} = 2\text{ bits}$$

Slide courtesy of Stephanie Forrest

Example: Horse race

- Probabilities of 8 horses are:

$$\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64}\right)$$

- Calculate the entropy H of a message containing the outcome:

- Suppose we want to send a msg saying who won
  - Could send the index of the winning horse (3 bits)
  - Or, use the following labels: 0, 10, 110, 1110, 111100, 111101, 111110, 111111
  - Avg. description length is 2 bits instead of 3

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{16}\log\frac{1}{16} - 4\frac{1}{64}\log\frac{1}{64} = 2 \text{ bits}$$

Slide courtesy of Stephanie Forrest

Another Example:

- Suppose you live in a city where ½ the time it's sunny and ½ the time it's cloudy. I say it's sunny. How much information did I give you?

- You live in Abq, 7/8 of the days its sunny, 1/8th it's cloudy. I say it's sunny. How much information did I give you?

Slide courtesy of Stephanie Forrest

Given: p(W = sunny) = 7/8 and p(W = cloudy) = 1/8

By observing W = sunny we get -log(7/8) ~=0.2 bits of info
By observing W = cloudy is -log(1/8) = 3 bits of info

On average, we observe W= sunny 7/8 of the time, and W = cloudy 1/8 of the time

The AVERAGE INFORMATION we receive by observing W is
-7/8*log(7/8) + -1/8 log(1/8) ~= 0.54
The SHANNON INFORMATION, $H(W_{abq})$ ~= 0.54

In New York half the days are sunny, and ½ the days are cloudy

The SHANNON INFORMATION, $H(W_{NY}) = 1$

Slide courtesy of Stephanie Forrest

Entropy and its Friends

- More generally,
  - The entropy of a random variable is a lower bound on the avg. number of bits required to represent the random variable
- The uncertainty (complexity) of a random variable can be extended to define the complexity of a single string
- e.g., Kolmogorov (algorithmic) complexity is the length of the shortest program that prints out the string.
- Entropy is the uncertainty of a single random variable
- Conditional entropy is the entropy of a random variable given another random variable

Slide courtesy of Stephanie Forrest

Mutual Information

- Measures the amount of information that one random variable contains about another random variable.
  - Mutual information is a measure of reduction of uncertainty due to another random variable.
  - That is, mutual information measures the dependence between two random variables.
  - It is symmetric in X and Y, and is always non-negative.
- Recall: Entropy of a random variable X is *H(X).*
- Conditional entropy of a random variable X given another random variable *Y = H(X | Y).*
- The *mutual information* of two random variables *X* and *Y is:*

$$I(X,Y) = H(X) - H(X|Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

This theorem, and the assumptions required for its proof, are in no way necessary for the present theory. It is given chiefly to lend a certain plausibility to some of our later definitions. The real justification of these definitions, however, will reside in their implications.

Quantities of the form $H = -\sum p_i \log p_i$ (the constant K merely amounts to a choice of a unit of measure) play a central role in information theory as measures of information, choice and uncertainty. The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics[8] where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the H in Boltzmann's famous H theorem. We shall call $H = -\sum p_i \log p_i$ the entropy of the set of probabilities $p_1, \ldots, p_n$. If x is a chance variable we will write $H(x)$ for its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from $H(y)$ say, the entropy of the chance variable y.

The entropy in the case of two possibilities with probabilities p and $q = 1 - p$, namely

$$H = -(p \log p + q \log q)$$

is plotted in Fig. 7 as a function of p.

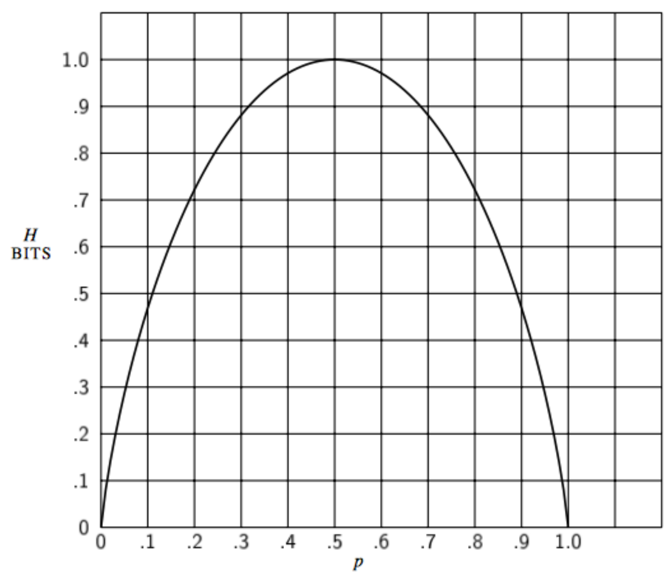
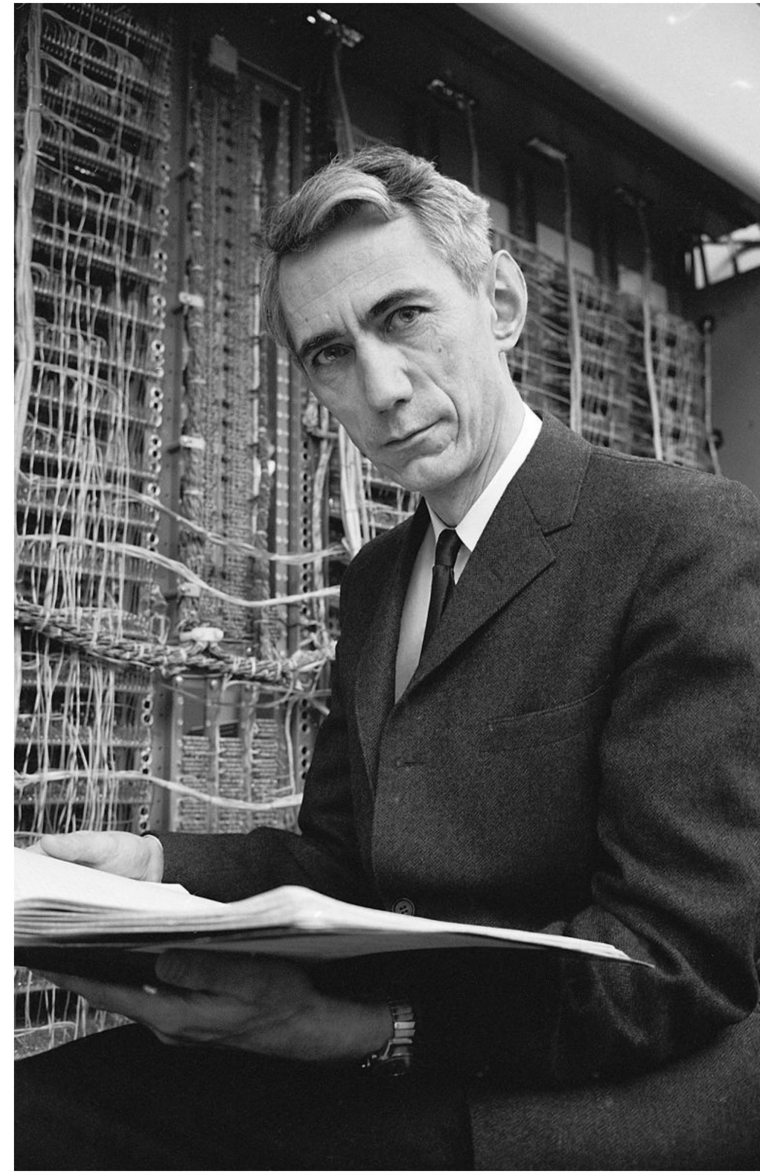


Fig. 7 — Entropy in the case of two possibilities with probabilities p and $(1 - p)$.

The quantity H has a number of interesting properties which further substantiate it as a reasonable measure of choice or information.

1. $H = 0$ if and only if all the p_i but one are zero, this one having the value unity. Thus only when we are certain of the outcome does H vanish. Otherwise H is positive.

2. For a given n, H is a maximum and equal to $\log n$ when all the p_i are equal (i.e., $\frac{1}{n}$). This is also intuitively the most uncertain situation.



Claude Shannon

A Mathematical Theory of Communication
The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948

Algorithmic Complexity (AC)
(also known as Kolmogorov-Chaitin complexity)

- The *Kolomogorov-Chaitin* complexity *K(x)* is the length, in bits, of the smallest program that when run on a Universal Turing Machine outputs (prints) *x* and then halts.
- Example: What is *K(x)* where *x* is the first 10 even natural numbers? Where *x* is the first 5 million even natural numbers?

Slide courtesy of Stephanie Forrest

Algorithmic Complexity (AC)
(also known as Kolmogorov-Chaitin complexity)

- Example: What is *K(x)* where *x* is the first 10 even natural numbers? Where *x* is the first 5 million even natural numbers?
- Possible representations:
  - *0, 2, 4, 6, 8, 10, 12, 14, 16, 18, … (2n - 2)*
  - for (j = 0; j < n: j++) printf("%d\n", j * 2);
- How many bits?
  - Alternative 1: *O(n log n)*
  - Alternative 2: *K(x) = O(log n)*
- Two problems:
  - Calculation of K(x) depends on the machine we have available (e.g., what if we have a machine with an instruction "print the first 10 even natural numbers"?)
  - In general, it is an uncomputable problem to determine K(x) for arbitrary x.

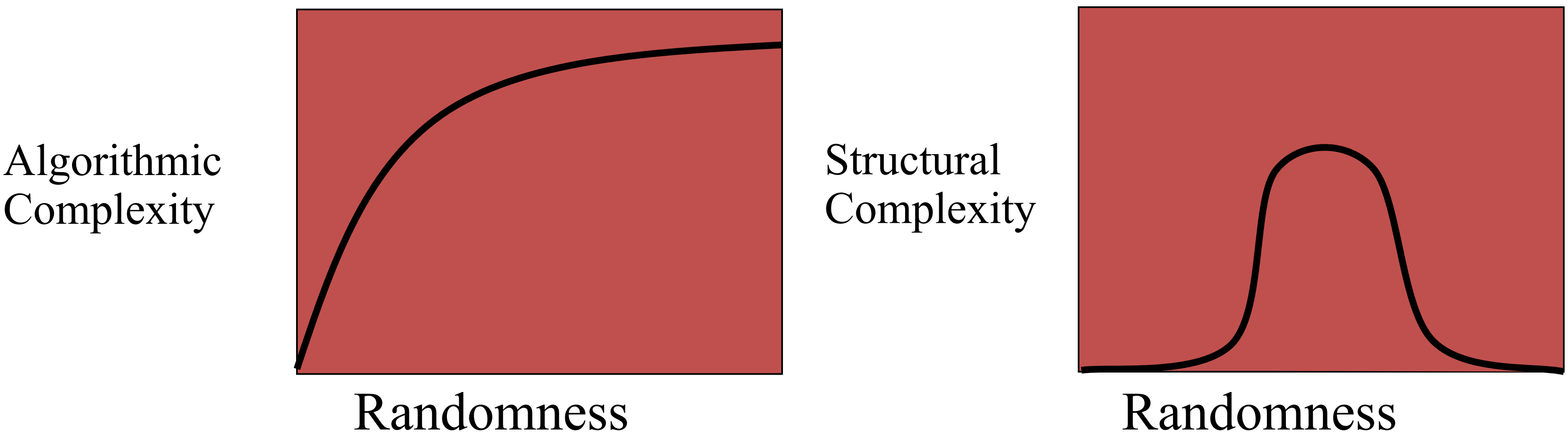
Slide courtesy of Stephanie Forrest

Algorithmic Complexity cont.

- AC formalizes what it means for a set of numbers to be compressible and incompressible.
  - Data that are redundant can be more easily described and have lower AC.
  - Data that have no clear pattern and no easy algorithmic description have high AC.
- What about random numbers? If a string is random, then it possesses no regularities:
  - *K(x)* = | Print(x) |
  - The shortest program to produce *x* is to input to the computer a copy of *x* and say "print this."
- Implication: The more random a system, the greater its AC.
- AC is related to entropy:
  - The *entropy rate* of a symbolic sequence measures the unpredictability (in bits per symbol) of the sequence.
  - The entropy rate is also known as the *entropy density* or the *metric density*.
  - The average growth rate of *K(x)* is equal to the entropy rate
  - For a sequence of n random variables, how does the entropy of the sequence grow with n?

Slide courtesy of [Stephanie Forrest]

Measures of Complexity that Capture Properties Distinct from Randomness



- Measures of randomness do not capture pattern, structure, correlation, or organization.
- Structural complexity
  - Mutual information, Wolfram's CA classification.
  - The "edge of chaos."

Slide courtesy of [Stephanie Forrest]

Logical Depth

- Bennett 1986;1990:
  - The *Logical depth* of *x* is the run time of the shortest program that will cause a UTM to produce *x* and then halt.
  - Logical depth is not a measure of randomness; it is small both for trivially ordered and random strings.
- Drawbacks:
  - Uncomputable.
  - Loses the ability to distinguish between systems that can be described by computational models less powerful than Turing Machines (e.g., finite-state machines).

Slide courtesy of Stephanie Forrest

A measure of complexity for Project 1

For each value of a used in the bifurcation plot, calculate the steady state entropy of x in the map. That is, after the map reaches the steady state, for each unique value of x_i calculate:

$$H_a(x) = -\sum_{x_i \in x} Pr[x_i] \log Pr[x_i] \tag{1}$$

What do you expect to see?

A measure of complexity for Project 1

For each value of a used in the bifurcation plot, calculate the steady state entropy of x in the map. That is, after the map reaches the steady state, for each unique value of x_i calculate:

$$H_a(x) = -\sum_{x_i \in x} Pr[x_i] \log Pr[x_i] \tag{1}$$

What do you expect to see?