CS 346U: Exploring Complexity in Science and Technology

Week 4 Assignment

Due October 28

I. Watch videos for Unit 4 on http://complexityexplorer.org **(due Wednesday October 23 by class time)**

II. Reading: Textbook, Chapter 3 **(due Wednesday October 23 by class time)**

III. Exercises (Ungraded and not turned in):

- All quizzes in Unit 4. Do these **before** you do the Unit 4 test.

IV. Unit 4 test (complete and submit online).

V. Assignments to turn in

1. Explain in your own words (one paragraph) how the second law of thermodynamics defines the ``arrow of time''.

2. Explain in your own words (one paragraph) the difference between "classical mechanics" and "statistical mechanics". (See Ch. 3 of the book.)

3. Explain in your own words (one paragraph) how Shannon information content is analogous to Boltzmann entropy.

4. In the Slot Machine example described in the lectures, calculate the number of microstates corresponding to:

(a) Exactly two of the same kind

(b) No lemons

(c) Two lemons and one orange

5. Suppose you have six fair dice, each with six sides.

The results when you roll all six dice at once is a microstate of the system. E.g., the microstate shown above is {6, 6, 6, 6, 6, 6}.

(a) Using $S(Macrostate) = \ln W$, what is S of the macrostate "each of the six dice shows the same number on its face"? Show your work.

(b) Same as (a) but for the macrostate "each of the six dice shows a different number on its face"? Show your work.

5. Suppose a one-year-old baby says five different words, each with equal probability. What is the Shannon information content of this one-year-old "message source"?

6. Suppose three-year-old Jake has a vocabulary of 500 words (including ``um"). When talking, he will say the word ``the" one-tenth of the time, the word ``um" one tenth of the time, and the rest of the time all his other words will be used equally often. What is the average Shannon information of his side of a conversation?

7. Experiment with TextInformationContent.nlogo, which can be downloaded from the Course Materials page. Use the model to see if information content is a good method for distinguishing, say, Spanish and English text, or any similar task. (Come up with your own tasks – be creative!) Describe your task and your results in two to three paragraphs.

8. Download LogisticMapInformationContent.nlogo from the Course Materials page. For each of R=2.0, 3.1, 3.49, 3.52, 4.0 (five different values), do the following:

Set x_0 to .2

Click "go", and let the model run for about 1000 ticks.

For each of these values of R, record the probabilities of 0 and 1, and the final value of information content. Do your own calculation of Shannon information content using these values and see if it agrees with the NetLogo model's results. Which values of R yields the highest information content, and why? Do you think this information content measure is a good measure of the complexity of the behavior of the logistic map? Why or why not? Describe your results and the answers to these questions in one or two paragraphs.

9. Modify LogisticMapInformationContent.nlogo to allow the user to set the threshold for the symbolic dynamics. (In the current version of the model, the threshold is hard-coded to 0.5 – that is, if $x_{t+1} < 0.5$, the system outputs "0", otherwise it outputs "1".) Investigate whether changing the threshold changes the information content measured for the values of R given in Assignment 8. That is:

- Set the threshold to 0.25 and repeat the steps of Assignment 8.
- Set the threshold to 0.75 and repeat the steps of Assignment 8.

Write a paragraph that describes your results and how the values of information content you obtained compares among the three thresholds (.25, .5, and .75).

10. Create a NetLogo model that simulates the rolling of a six-sided die and calculates the Shannon information content of the results. I.e., the user rolls the die N times. Assume that the probability of each "message" (number from 1 to 6) is its relative frequency over these rolls. Use these probabilities to calculate Shannon information content. (We will work on this in class together.)

11. Optional (for more advanced programmers): Read about "Huffman Coding" (e.g.,

http://www.cse.ust.hk/faculty/golin/COMP271Sp03/Notes/MyL17.pdf

or

http://www.cs.duke.edu/csed/poop/huff/info/

Implement a program in NetLogo (or whatever language you prefer) to input a text and to output a Huffman code for that text. Does the average number of bits per "message" you get match the information content of the text?

What to turn in: Complete and submit the Unit 4 test online. Email a pdf document with your typed answers to Assignments 1-9, along with your NetLogo model from Assignment 10, to mm@cs.pdx.edu. If you prefer, you can write the math parts by hand and turn them in to me in class on Monday.