Introduction to Complexity

Unit 4 Homework

(optional)

Beginner Level:

1. 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

2. 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"? (Note that "ln" is the notation for "natural logarithm".)

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

3. Consider the six-dice system as a "message source" and a roll of the dice as a "message". If all the dice are fair (equal probability for each value), what is the Shannon information content of this message source?

4. 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"?

5. Come up with a real-world example of a "message source" for which you could calculate Shannon information content. What is the Shannon information content of your example message source?

6. 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.

7. 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?

Intermediate Level:

1. 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 sixth 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?

2. Modify SlotMachine.nlogo to include the following macrostates (put these in the drop-down menu for Macrostates):

(a) Exactly two of the same kind

(b) No lemons

(c) Two lemons and one orange

3. Modify CoinFlipInformationContent.nlogo to be "DiceRollInformationContent.nlogo" where, instead of a coin flip, a six-sided die is rolled.

4. 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 various values of R.

Advanced Level:

1. 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?