Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Department of Mechanical Engineering

6.050J/2.110J **Information and Entropy** **Spring 2004**

Issued: March 29, 2004

Problem Set 8

Due: April 2, 2004

Note: The **quiz** will be held Monday, April 12, 2004, 1:00 PM - 2:00 PM, in Room 2-105. The quiz will be **closed book** except that you may bring one sheet of 8 1/2 x 11 inch paper with notes on both sides. Calculators will not be necessary but you may bring one if you wish. Material through the end of Problem Set 8 may be covered on the quiz.

Problem 1: Uncertain Employment

You have graduated from MIT with high honors, and have moved on to the field of politics, your passion. After many years of hard work, you eventually become advisor to the President of the United States on employment policy. According to the classification system you have devised there are three possibilities for the employment status of a person:

- Fully employed (100% of capacity)
- Semi-employed (50% of capacity)
- Unemployed (0% capacity)

The President is trying to set economic policy so that the employment rate of the country is kept to 90%, as measured with your system (i.e., each half-employed person counts as half a fully-employed person) You do not know the exact statistics yet of the economic data, and so you express what knowledge you do have in terms of probabilities – the probabilities $p(E)$ (fully-employed), $p(S)$ (semi-employed), and $p(U)$ (unemployed) of the employment status of any particular citizen. Since each citizen falls into exactly one of the categories, you know that the events E, S, and U form a partition, so

$$p(E) + p(S) + p(U) = 1 \tag{8–1}$$

Furthermore, you know that when the policy goal is achieved, the average employment will be 0.9:

$$1.0p(E) + 0.5p(S) + 0.0p(U) = 0.9 \tag{8–2}$$

Noticing that there are many possible probability distributions consistent with these two constraints, you decide to find the distribution that uses only the information you have. In other words, you use the principle of maximum entropy. (This approach is consistent with your experience that usually politicians reveal the minimum amount of information about the the success of their policies.)

Note: To do this problem, you may want to use MATLAB. If you do and then hand in your assignment on paper, please write the MATLAB statement you used. Remember to show all work. Note that `log2(x)` is not implemented correctly on all MATLAB versions, so use `log(x)/log(2)` instead. The commands `solve` and `diff` can be used to find where a function's derivative is zero; to see the help files for these operations, type `help sym/solve` or `help sym/diff`.

a. Recall that all probabilities must lie in the interval between 0 and 1.

   i. What range of $p(E)$, the probability that any one citizen is fully-employed, is consistent with these constraints?

   ii. Plot the entropy of the probability distribution over this range as a function of $p(E)$. (If you turn in the problem set on paper, be sure to include this graph!)

   iii. Find the maximum entropy, and give the value of all three probabilities at this maximum.

b. An election is coming up, and the President's campaign manager believes that unemployed and semi-employed people would be apt to vote for the opposition. Therefore he would like to see a probability distribution that, while being consistent with the employment goal, would have the largest number of fully employed people. Without calculating it, say whether this probability distribution will be less than, equal to, or greater than the valued you calculated in part (a-iii) above.

c. Naturally you wonder what exactly the entropy of this distribution will be. Find the campaign manager's preferred distribution (all three probabilities) and the entropy.

d. The President's speech writer wants to put the employment policy in the best possible light, so she asks for a probability distribution that has the least number of unemployed people but is consistent with the employment goal. What is this probability distribution and what is its entropy?

e. The President is not sure what entropy is, but hears that it is a measure of uncertainty. He does not want his administration to have an image of uncertainty, so he asks for a probability distribution with minimum, rather than maximum entropy. He is willing to change his employment goal if necessary. You are asked for the probability distribution that has the least possible entropy. What is that value of entropy? Give at least one distribution that achieves that value.

Problem 2: Candy Raffle

You have entered a raffle at the Cambridgeside Galleria wherein you must guess certain properties of a jar filled with Tootsie Rolls, M&M's, and Hershey's Kisses. Normally you would guess only how many pieces of candy were in the jar, but the Galleria is wise to the ways of MIT students, and so has decided to make the raffle slightly more difficult than normal by asking for properties other than the number of pieces of candy. You know that M&M's weigh 0.8 grams, Tootsie Rolls weigh 2.5 grams, and a Hershey's Kiss weighs 1.4 grams. You also know that Tootsie Rolls are 50% sugar, M&M's are 64% sugar, and Hershey's Kisses are 8% sugar. (This information can be gathered from the Hershey's, M&M, and Tootsie Roll websites). The candy in the jar has a total mass of 2400 grams, but you don't know how many of each type are in the jar. To win the raffle, you are asked to provide the answers to the following questions:

a. What ranges of numbers for each type of candy in the jar are compatible with this total mass?

b. You are told there are 1000 pieces of candy in the jar. Make no additional assumptions. Determine those numbers of each type of candy that maximize your remaining uncertainty. Don't worry if you get fractional pieces of candy.

c. Given the numbers you calculated, what do you expect the percentage by mass of sugar inside the jar to be?

d. Can you have 600 Tootsie Rolls? Is this consistent with the answer in part (a)? Explain.

Turning in Your Solutions

If you used MATLAB for this problem set, you may have some M-files and a diary. Name the M-files with names like `ps8p1.m`, `ps8p2.m`, and name the diary `ps8diary`. You may turn in this problem set by e-mailing your written solutions, M-files, and diary to 6.050-submit@mit.edu. Do this either by attaching them to the e-mail as *text* files, or by pasting their content directly into the body of the e-mail (if you do the latter, please indicate clearly where each file begins and ends). If you have figures or diagrams you may include them as graphics files (GIF, JPG or PDF preferred) attached to your email. Alternatively, you may turn in your solutions on paper in room 38-344. The deadline for submission is the same no matter which option you choose.

Your solutions are due 5:00 PM on Friday, April 2, 2004. Later that day, solutions will be posted on the course website.