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

**6.050J/2.110J** **Information and Entropy** **Spring 2005**

**Issued: March 28, 2005** **Problem Set 8** **Due: April 1, 2005**

**Note:** The **quiz** will be held Tuesday, April 12, 2005, 12:00 PM - 1:00 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 (including children and those who have retired):

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

The President is trying to set economic policy so that the employment rate of the country is kept to 60%, as measured with your system (i.e., each semi-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 0.60:

$$1.0p(E) + 0.5p(S) + 0.0p(U) = 0.60 \tag{8–2}$$

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, and the resulting employment goal.

---

## Problem 2: Lost in Translation

You and two colleagues (an Englishman and a Spaniard) want to design the perfect vehicle of the future—an environmentally friendly rocket with an anti-gravity engine. You know that to convince venture capitalists to give you money, you will need a very well written, inspiring proposal. Your two colleagues have written drafts, which were edited by an English major from Harvard. You are in charge of the final edit.

You look at the three verbs "fly," "sail," and "glide," which occur frequently. The Harvard student did not appreciate the subtle technical differences in the meaning of those words, and to enliven the text he substituted one for another randomly. In particular, when the original draft used the word "fly" the student admitted he replaced it by "sail" one third of the time and "glide" one third of the time. Similarly, "sail" was replaced by "fly" 30% of the time, but never by "'glide." Finally, "glide" was replaced by "sail" 10% of the time but never by "fly." To correct these random substitutions, you want to infer what the author meant in each case. That is, given one of the three words in the edited text, you want to infer the original word used by the author.

a. Your great knowledge of Information and Entropy allows you to make these inferences. First, represent the editor by a Probability Diagram with inputs $F$, $S$, and $G$ for the three words, and outputs $f$, $s$, and $g$.

You will develop rules to say which input was most likely given each of the three possible outputs.

b. Your accent-driven bias makes you assume that your Spanish colleague knows less English and used the most common verb "fly," 90% of the time, "sail" 10%, and "glide" never. With this assumption of input probabilities, find, for each of the three output verbs, the word most likely to have been used by the author.

c. Next, biased by your admiration for British elegance, you assume (again incorrectly) that your British colleague wanted to avoid "fly" and guessed he used it only 10% of the time, and "glide" 40%. With this assumption of input probabilities, find, for each of the three output verbs, the word most likely to have been used by the author.

When your colleagues see your reconstruction of the meaning they intended for each sentence they let you know that you got them all wrong. Realizing that you have no reason to believe either of these probability distributions, you know it is time to use another principle besides your own prejudices. Naturally, to avoid unintended bias, you use the Principle of Maximum Entropy.

d. Write an expression for the uncertainty of the word used, as a function of the three probabilities $p(F)$, $p(S)$, and $p(G)$, with the sole constraint that they add up to one. What are the values of these three quantitites that give a maximum uncertainty, and what is the uncertainty in bits?

You speak with the authors and discover that they tried to adhere to a definite degree of elegance. They assigned an "elegance value" to each of the words: $e(S)$ was set to $-10$ because it might imply (incorrectly) the need for wind; $e(G)$ was set to $+10$ because of its connotation of gracefulness, and $e(F)$ was set to 0. You discover that your earlier assumptions both implied an average elegance of $-1$ and so you decide to look for the maximally unbiased probability distribution with an expected value of elegance of $-1$.

e. What probability distribution maximizes the uncertainty, consistent with this assumed average elegance? What is the resulting uncertainty, in bits?

f. We say that the choice of a-priori probability distribution from parts b and c introduces bias, while the maximum entropy principle does not. Nevertheless, albeit unbiased, the a-priori probability distribution resulting from the application of the maximum entropy principle is still subjective. Explain briefly bias and subjectiveness in this context.

---

# 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 1, 2005. Later that day, solutions will be posted on the course website.