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 15, 2004

Problem Set 7

Due: March 19, 2004

Problem 1: Communications Crosstalk

The process model covered this week can be used for both deterministic systems, whose output is determined by the input, and by nondeterministic systems. Let's use it to describe the action of a a communications line which has crosstalk. The communications channel is show in Figure 7–1. The channel is intended to carry a 1 or 0 to a distant location (this is represented by the first bit of each input bit pair) . There is an additional factor, the charge on a nearby line, which we will model as the second bit of each input bit pair. That is, when the nearby line has a 1, and the communications channel we're interested in has a 1, this input is 11. Similarly, if the nearby line has a 0 instead, this is represented as a 10.

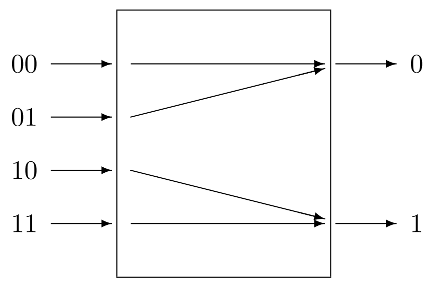

Figure 7–1: A communications channel

a. First, consider a channel without crosstalk, as shown in Figure 7–1. Assume each of the four possible inputs is equally likely, for example if the input had been obtained by two independent coin tosses. Calculate the two output probabilities $p(B_0)$ and $p(B_1)$, the input information I_{in}, the output information I_{out}, the noise N, loss L, and mutual information M.

b. Now consider the effects of crosstalk. Assume that whenever the second bit is high (i.e., the nearby line has a value of 1) there is a 20% chance that the first bit is flipped during communication.

   Draw a process model diagram which models the communication crosstalk as a process. Include the transition probabilities in your diagram.

c. If the output is 1. . .

   i. What is the probability that it was produced by the input (0 1)?
   ii. What is the probability that it was produced by the input (1 0)?
   iii. What is the probability that it was produced by the input (1 1)?
   iv. What is the probability that it was produced by the input (0 0)?

d. What are the input information I_{in} and the output information I_{out} (in the correct units)?

e. What are the noise N, the loss L, and the mutual information M? Is this process noisy, lossy, both, or neither?

f. **Extra Credit:** How useful is the comparison between I_{out} for the correct channel and I_{out} for the channel with crosstalk? What about the comparison between M for the two channels? In answering this question, you might consider whether crosstalk would be a useful feature in such a communications line.

Problem 2: Cantabridgian Computer Competency

The City of Cambridge has recently passed an ordinance requiring all freshmen in Cambridge to pass a computer-literacy test. For simplicity we will consider that there are only two schools, Harvard (hereafter to be symbolized by the letter 'H') and MIT (M). The result of the exam is such that each freshman is deemed Competent (C) or Incompetent (I). This will be called the 'competency' of the student. Note that H/M form a partition and C/I form another partition of the set of all freshmen.

It is known before the exam that every MIT freshman is competent with computers, but only one-third of the Harvard freshmen are competent. After the exam has been administered to all freshmen at Harvard and MIT, the City would like to tabulate the results by school. Unfortunately, the proctor of the exam forgot to ask each freshman to mark down their schools, so there is no direct way to tell if a particular freshman was from Harvard or MIT.

For simplicity, assume that Harvard has 3,000 freshmen and MIT 1,000 and that all the freshmen have taken the exam.

a. What is the probability $p(M)$ that a freshman selected at random is from MIT?

b. Since you are currently taking 6.050/2.110 at MIT, you decide to model the survey as a non-deterministic process, where the input is the school (H/M) and the output is test result (C/I). Then guessing the school is similar to what is done in communication channels, where the input is inferred from knowing the output. Now assume that a freshman is chosen at random but you do not know if she is competent. What is your uncertainty (measured in bits to two decimal places) of that student's school?

c. What is your uncertainty in school if you are told that the freshman was deemed competent?

d. What is your uncertainty if you discover that the freshman was incompetent?

e. What is the uncertainty, on average, of school once you learn the competency of the freshman?

f. City Hall needs to correlate survey results with school, but doesn't want to conduct another survey. To help them, you decide to design a probabilistic inference machine to guess H/M given C/I. Your design is quite simple: if the probability of a freshman being from MIT given that the student is competent $p(M \mid C)$ were 3/4 (clearly it's not), then your machine would guess M with probability 3/4 every time you give it C. Your inference machine can be modelled as a process. Design the inference machine. In other words, give $p(H \mid C)$ (the probability of H given C), $p(H \mid I)$, etc. Present your answer in the form of a process model diagram, like the ones discussed in lecture and in the notes.

g. Like a true nerd, you begin to think of the combination of the exam and your inference machine as a binary communications channel (this is shown in Figure 7–2). The "input" to this channel is the school of a student selected at random, and the "output" is your machine's guess about the school. Naturally you wonder about the amount of information such a channel could pass. Compute the input and output information, and the noise, loss, and mutual information of the combined system. As usual, your results should be in the proper units, rounded to two decimal places.

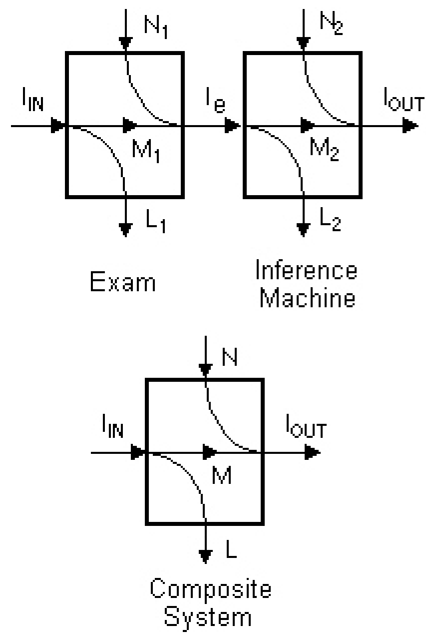

Figure 7–2: Example process model combining exam and inference machine

h. Verify one of the four inequalities for the mutual information M in terms of M_1 and M_2 at the end of Chapter 7 of the notes, where M_1 and M_2 are each the mutual information of the first (exam) and second (inference machine) subparts of the whole communication channel.

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 `ps7p1.m`, `ps7p2.m`, and name the diary `ps7diary`. 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, March 19, 2004. Later that day, solutions will be posted on the course website.