6.050J/2.110J	Information and Entropy	Spring 2003

Issued: March 14, 2003   # Problem Set 6   **Due: March 21, 2003**

Problem 1: What Kind of Gate is This?

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 peculiar gate that detects whether one input is greater than the other. A properly working gate of this type has the model show in Figure 6–1.
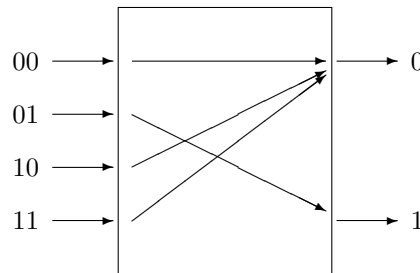

Figure 6–1: The "Greater" gate

a. 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. You have purchased some of the gates from Mark's Warehouse, which has low prices but no refunds. The order arrives with a note saying that, due to a "slight manufacturing error," your gates do not behave exactly as they should. Specifically:

- the outputs for inputs (1 1) and (0 0) are correct.
- the output for input (0 1) will be wrong 10% of the time.
- the output for input (1 0) will be wrong 20% of the time.

Since you cannot return them and you don't have time to buy new gates before your project is due, you decide to make the best of things, and start by analyzing these defective gates. Continue to assume that all four inputs are equally probable.

Draw a box diagram which models the defective gate 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)?

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, or both?

f. **Extra Credit:** How useful is the comparison between I_{out} for the correct gate and I_{out} for the defective gate? What about the comparison between M for the two gates? In answering this question, you might consider which kind of gate would, in general, be more useful in circuit design.

Problem 2: Cantabridgian Computer Competency

The City of Cambridge has recently passed an ordinance requiring all students 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 student 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.

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

    For simplicity, assume that Harvard has 20,000 students and MIT 10,000 and that all the students have taken the exam.

a. What is the probability $p(M)$ that any given student who took the exam is from MIT?

b. Having aced 6.050/2.110 as a freshman back at MIT, you decide to model the survey as a nondeterministic 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. What is the uncertainty (measured in bits to two decimal places) of school if you do not yet know the competency of a random student who took the exam?

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

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

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

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 student 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 box diagram, like the ones discussed in lecture and in the notes.

g. Suddenly you realize that you can think of the combination of the exam and your inference machine as a binary communications channel: a student taking the exam "sends" his or her school information through this channel by specifying only their competency. You wonder how good a communications channel this might be. Evaluate the performance of this channel by computing the noise, loss, and mutual information of data carried across the combined system. As usual, your answer should be in bits rounded to two decimal places.

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 `ps6p1.m`, `ps6p2.m`, and name the diary `ps6diary`. 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 21, 2003. Later that day, solutions will be posted on the course website.

Enjoy Spring Vacation! We will see you again April 1, no fooling!