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 14, 2005 **Problem Set 7** **Due: March 18, 2005**

Problem 1: Confusius Gate...

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 peculiar gate. The gate in Figure 7–1 will output the opposite of what the input was when both input bits are the same, and the second input bit if the two bits in the input differ.

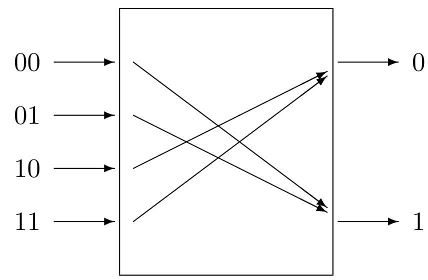

Figure 7–1: A confusius gate

a. First, consider the channel without defects, 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, the output information J, the noise N, loss L, and mutual information M.

b. Lupus Berevus, the only manufacturer of these gates in the world, happens to be dyslexic (that's how he came up with the gate to start with), and is unable to remember the rules for 01 and 10 which he keeps on crossing while he constructs the gates. In average, about a 40% of the time the output of 10 is changed by that of 01 and vice versa. In addition to that the gates are not properly shielded from external currents and this affects the output for input 00 that ends up being interpreted as any other input with equal probability.

   Draw a process model diagram which models the confusing confusius 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)?
   iv. What is the probability that it was produced by the input (0 0)?

d. What are the input information I and the output information J (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 J for the correct channel and J for the channel with defects? What about the comparison between M for the two channels? In answering this question, you might consider thinking about which gate may be more useful in circuit design.

Problem 2: And then came the humans, ...

According to the latest archeological findings, about two hundred thousand years ago, two species of hominids *Homo Sapiens*, and *Homo neanderthalis*, coexisted with the currently dominant *Homo sapiens sapiens* for as long as one hundred thousand years. It has been debated what is the exact relationship between the three species in terms of evolution. In this problem set we will frame this debate as an information flow problem.

Let us first focus on a hypothetical interbreeding between *H. sapiens* (which we abbreviate as **S**), and *H. neanderthalis* (which we abbreviate as **N**). For the moment we will ignore mutations leading to *H. sapiens sapiens*. We will address this problem taking the mothers of each species as input. Therefore, we will have two possible values as input; either the mother is "S" and we label the input M_S or she is "N" and we label the input M_N. Assume a female population of 1000 N and 10000 S.

In order to characterize the evolution process, we need to know the number of females that have children, and the species of the children. We do not really have observations from that time, so we will have to guess; the following assumptions (guesses) will help us characterize the system. Assume that females mate only once. Typically females from one species will mate males from the same species. When that is the case, the females will always have children (labeled C_S or C_N according to the species of the parents). However, some interspecies mating occurs. Suppose 20% of N-females will mate S-males, and 10% of S-females will mate N-males. Fertility of interspecies couples is reduced; only 50% of them have children. Concerning the species of the children of interspecies couples, 10% will be classified as C_N, and the remaining 90% will be seen as C_S.

a. What is the probability that a randomly chosen mother is *H. neanderthalis* ($P(M_N)$)? (HINT: Only females with offspring are mothers!. You should get $P(M_N) = .0865$)

b. As an observer *H. sapiens sapiens* that has forgotten his glasses, you see a kid playing around but your miopy does not let you appreciate to which species he belongs. What is your uncertainty (measured in bits up to two decimal places) about the species of the mother?

c. Model evolution as a nondeterministic process. (We will later refer to this model as "expected evolution"). The mothers of each species are the input of the system, and their offspring are the output of the system. (This is a probability model, like the ones you have seen in class). Draw the diagram of the probability model and compute its transition probabilities. (HINT: Two of the transition probabilities you should get are $C_{SN} = 1/10$ and $C_{NS} = 0.0053$)

d. Compute $P(C_N)$ and $P(C_S)$.

e. If you are told that the kid playing around appears to be S, what is your uncertainty about the species of the mother?

f. What if you were told that the kid is N?

g. You can set up an inference machine (one that tells you what is the probability of the mother being S or N given the offspring's species.) Present this machine in the form of a probability model diagram like the ones discussed in class.

Now, you will complete the model of evolution by introducing mutations, so that the model also includes *H. sapiens sapiens* (which we abbreviate as **H**). We will do so by introducing a second process that we will call "Mutation". The input of the mutation process is the output (C_N, C_S) of the "expected evolution" process that you defined in part c; the output of the mutation process is the actual result of evolution once mutation is taken into account. Denote the outputs of the process of mutation as J_N, J_S, J_H. [10]

Assume that *H. sapiens sapiens* are mutants of *H. sapiens*, and that mutations occur 1% of the time.

h. Characterize the probability model of the process of mutation the same way you did for the model of "expected evolution".

i. You are ready to examine the entire model of evolution including mutations. Combine the "expected evolution" model with the mutation model as a cascade forming a single process. Compute the input and output information and the noise and the loss of the combined system. (Make sure you draw the corresponding process diagrams)

j. **EXTRA CREDIT** Verify one of the four inequalities, equations 7.37-7.40, for the mutual information M in terms of M_1 and M_2 in Chapter 7 of the notes, where M_1 and M_2 are each the mutual information of the first ("Expected Evolution") and second ("Mutation") 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 18, 2005. Later that day, solutions will be posted on the course website.

[10] Do note get confused by this model. The model does not imply that mutations occur at the offspring; they really occur at the parent level. Our model simply compares expected outcome with actual outcome when mutations are factored in. Unlike many Sci-Fi movies make us believe, individuals do not turn into mutants, they may be born so with respect to their parents.