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 7, 2005** **Due: March 11, 2005**

Problem Set 6 Solutions

Solution to Problem 1: When Hell Freezes Over..

Solution to Problem 1, part a.

The probability of MIT closing on any given day, can be found as follows:

$$P(C) = \frac{P(C|F) * P(F)}{P(F|C)} \tag{6–1}$$
$$= \frac{0.15 * 0.01}{1} \tag{6–2}$$
$$= 0.0015 \tag{6–3}$$

Solution to Problem 1, part b.

The probability of MIT closing and Hell freezing over on the same day.

$$P(C, F) = P(C|F) \times P(F) \tag{6–4}$$
$$= 0.15 * 0.01 \tag{6–5}$$
$$= 0.0015 \tag{6–6}$$

OR

$$P(C, F) = P(F|C) \times P(C) \tag{6–7}$$
$$= 1 * 0.0015 \tag{6–8}$$
$$= 0.0015 \tag{6–9}$$

(Note that there are multiple ways to represent the joint probability symbolically. Acceptable forms are $P(C, F)$, $P(C\&F)$, and $P(CF)$. The order of C and F is unimportant.)

Solution to Problem 1, part c.

Hell is expected to freeze over:

$$365 * P(F) = 3.65 \text{ days a year.}$$

Solution to Problem 1, part d.

MIT is expected to close:

$$365 * P(C) = 0.5475 \text{ days a year.}$$

Solution to Problem 1, part e.

In general to answer this question we would need to compute $P(C, 2006|C, 2005)$. That can be readily done using Bayes' rule: $P(C, 2006|C, 2005) = P(C, 2006, C, 2005)/P(C, 2005)$. Unfortunately we do not have a model of the joint probability for MIT closing in two different years. In these conditions a common assumption is that both events are independent (much like coin tosses are), in which case the answer is simply:

$$P(C, 2006|C, 2005) = P(C, 2006, C, 2005)/P(C, 2005) \tag{6–10}$$

$$= P(C, 2006)P(C, 2005)/P(C, 2005) \tag{6–11}$$

$$= P(C, 2006) \tag{6–12}$$

$$\tag{6–13}$$

and we can safely take $P(C, 2006) = P(C, 2005)$.

Solution to Problem 2: Communicate

Solution to Problem 2, part a.

Since this is a symmetric binary channel (equal probability of a 1 switching to a 0 as does a 0 switch to a 1), we can use the following derivative of Shannon's channel capacity formula. H(x) is defined to be the number of bits of information in x. Note that x is not an argument of H. You can read that notation as "The entropy of x" or "The amount of unknown information in x".

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{6–14}$$

$$H(transmitted) = H(input) - H(loss)$$

$$W = Channel\ Bandwidth = (Max\ Channel\ capacity\ with\ errors)$$

$$C = Max\ Channel\ capacity\ (without\ errors) = W * H(transmitted)$$

In these equations the loss term is the information lost when the physical noise changes the bits within the channel. We assume that the input bitstream has equal probability of 1's as 0's, so H(input) = 1. We assume this because otherwise we wouldn't be utilizing all of the channel bandwidth. So, our next task is to determine the number of bits of information lost because of errors in the channel. This is the information we do not have about the input even if we know the output.

$$H(loss) = -0.25 * \log_2(0.25) - 0.75 * \log_2(0.75) \tag{6–15}$$

$$= 0.5 + 0.3113 \tag{6–16}$$

$$H(loss) = 0.8113 \tag{6–17}$$

$$\tag{6–18}$$

$H(loss)$ is the number of bits that are distorted per input bit. The difference with th einitial entropy or uncertainty is the amount of information that makes it to the output uncorrupted: *the information transmitted*. Applying these results yields the following.

$$H(transmitted) = 1 - 0.8113 = 0.1887 \tag{6–19}$$

$$W = 1000 bits/sec \tag{6–20}$$

$$C = 1000 * 0.1887 \tag{6–21}$$

$$= 188.7 bits/sec \tag{6–22}$$

Solution to Problem 2, part b.

Since triple redundancy is being used, we'll encapsulate the source encoder to form a new channel as in the following diagram.

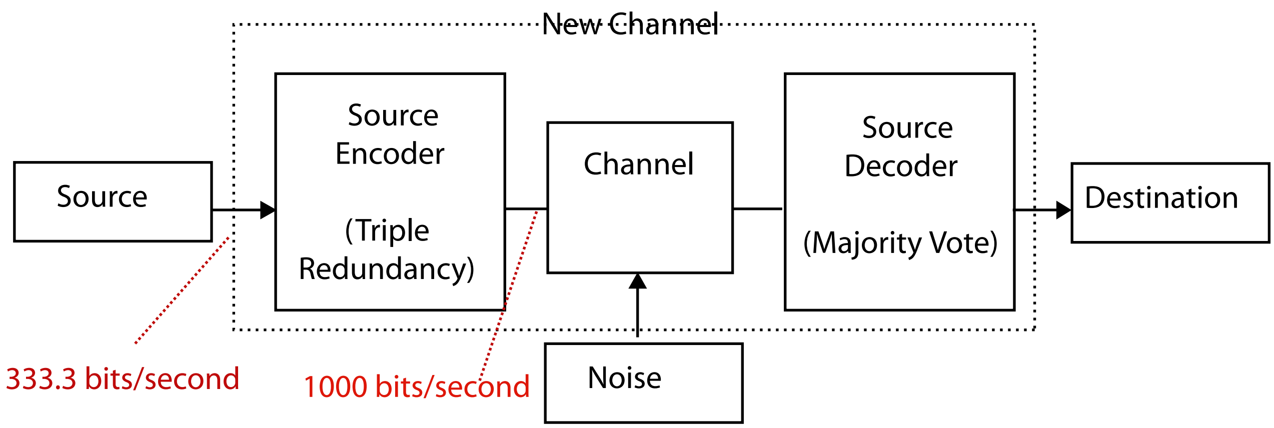

Figure 6–1: New channel model after addition of triple redundancy

Since the original channel has a bandwidth of 1000 bit/sec and we are now using triple redundancy, our new maximum channel capacity that we can feed into our new channel is 333 1/3 bits/sec.

Solution to Problem 2, part c.

Also, our channel characteristics change. No longer is the probability of a bit error 25%. Using Bayes theorem and the binomial distribution to establish theprobability of having a given number of errors, we can obtain the new probability of a bit error. Let A be the event that redundancy fails, and let us call ε the original error rate.

$$P(A, 0 \text{ errors}) = P(A|0 \text{ errors}) \times P(0 \text{ errors}) = 0 \quad (6\text{–}23)$$
$$P(A, 1 \text{ error}) = P(A|1 \text{ error}) \times P(1 \text{ error}) = 0 \quad (6\text{–}24)$$
$$P(A, 2 \text{ errors}) = P(A|2 \text{ errors}) \times P(2 \text{ errors}) \quad (6\text{–}25)$$
$$= 1 \times \binom{3}{2} \varepsilon^2 (1-\varepsilon)^1 \quad (6\text{–}26)$$
$$= 1 \times 3 * (0.25)^2 * (1 - 0.25)) \quad (6\text{–}27)$$
$$P(A, 3 \text{ errors}) = P(A|3 \text{ errors}) * P(3 \text{ errors}) = \quad (6\text{–}28)$$
$$= 1 \times \binom{3}{3} \varepsilon^3 \quad (6\text{–}29)$$
$$= 1 \times (0.25)^3 \quad (6\text{–}30)$$
$$P(A) = P(A, 0 \text{ errors}) + P(A, 1 \text{ error}) + P(A, 2 \text{ errors}) + P(A, 3 \text{ errors}) \quad (6\text{–}31)$$
$$P(A) = 0.1406 + 0.0156 \quad (6\text{–}32)$$
$$P(A) = 0.1562 \quad (6\text{–}33)$$

Solution to Problem 2, part d.

Continuing on with our calculations to determine the entropy of the loss and the resulting maximum channel capacity without errors yields the following.

$$H(loss) = -0.1562 * \log_2(0.1562) - 0.8438 * \log_2(0.8438) \quad (6\text{–}34)$$
$$H(loss) = 0.4184 + 0.2068 \quad (6\text{–}35)$$
$$H(loss) = 0.6251 \quad (6\text{–}36)$$
$$(6\text{–}37)$$
$$H(transmitted) = 1 - 0.6251 = 0.3749 \quad (6\text{–}38)$$
$$W = 333.333 bits/sec \quad (6\text{–}39)$$
$$C = 333.333 * 0.3749 = 124.95 bits/sec \quad (6\text{–}40)$$

This new maximum channel capacity is less than that of the original channel. From this result, we can see the tradeoff for decreased bit error probability for bandwidth and the resulting maximum channel capacity.

Solution to Problem 2, part e.

Our probability of bit error in this new channel is above the 10% required for the application being designed for. However, the probability of error is indeed less than what it was before. The price we pay is a lower bitrate.

Solution to Problem 3: Nerds' Nest Expands Operations

The first thing to do in this problem set is to identify what type of communication does the extravagant set-up correspond to. Figure 6–2, shows how to do that.

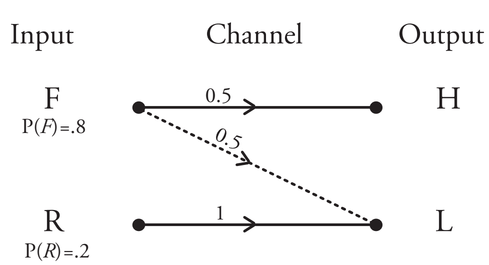

Figure 6–2: Scheme for Nerd's Nest entree communication system, transition probabilities are expressed on top of the lines connecting input with output.

The conclusion is that we are dealing with an instance of an asymmetric binary communication channel. We can just ignore the rest of the detail and characterize this communication system with section 6.6 of the notes ("Noisy Channel") as reference.

Solution to Problem 3, part a.

The first difference with the course notes is that our communication channel, is not symmetric, therefore, we can expect to have a different transition matrix:

$$\begin{bmatrix} c_{HF} & c_{HR} \\ c_{LF} & c_{LR} \end{bmatrix} = \begin{bmatrix} .5 & 0 \\ .5 & 1 \end{bmatrix} \tag{6–41}$$

Solution to Problem 3, part b.

The probabilities of the outcomes can be obtained by Bayes' rule, and summing over all possible inputs:

$$P(H) = P(H|F)P(F) + P(H|R)P(R) \quad = c_{HF}P(F) + c_{HR}P(R) = .5 \times .8 + 0 \times .2P(H) \quad = .4 \tag{6–42}$$

$$P(L) = P(L|F)P(F) + P(L|R)P(R) \quad = c_{LF}P(F) + c_{LR}P(R) = .5 \times .8 + 1 \times .2P(L) \quad = .6 \tag{6–43}$$

Solution to Problem 3, part c.

In order to assess the uncertainty before transmission occurs, we need to look at the input probabilities, that were given in the problem statement: $P(R) = .2 = 1 - P(F)$. Then according to the lecture notes the uncertainty before transmission is (also known as entropy and often referred to with the letter H):

$$U_{\text{before}} = P(R) \log_. 2 \left(\frac{1}{P(R)} \right) + P(F) \log_2 \left(\frac{1}{P(F)} \right) \tag{6–44}$$

$$= .4644 + .2575 \tag{6–45}$$

$$U_{\text{before}} = .7219 \tag{6–46}$$

Solution to Problem 3, part d.

In the absence of error, the capacity of a communicaton channel is expressed as:

$$C = W \log_2 n \tag{6–47}$$

where n is number of possible input states, and W is the transmission rate (in bits per second). In the case of a binary channel $n = 2$, and the problem statement tells us that $W = 1$ orders/second. So the ideal (noise-free) capacity of the system is:

$$C = 1bit/second$$

.

You may wonder why we did not use the input probabilities; if so, remember that Capacity is a property of the channel. You could have taken the capacity to be defined as $C = W \times M_{max}$, as it is for the noisy case, even with this definition you must find the same result because $M_{max} = 1$ in the symmetric binary channel.

Solution to Problem 3, part e.

Back to the noisy case, we want to compute the mutual information between input and output. In the solution to previous problems we have referred to it as $H(transmitted)$. It is the same concept. The lecture notes define this quantity in equations (6.25 & 6.26). The easiest here is to take the second definition in equation (6.26) because we have already computed all the relevant intermediate quantities.

$$M = \sum_j P(B_j) \log_2 \left(\frac{1}{B_j} \right) - \sum_i P(A_i) \sum_j P(B_j|A_i) \log_2 \left(\frac{1}{P(B_j|A_i)} \right) \tag{6–48}$$

adapting it to our problem

$$\tag{6–49}$$

$$=P(L)\log_2\left(\frac{1}{P(L)}\right)+P(H)\log_2\left(\frac{1}{P(H)}\right) \tag{6–50}$$

$$-P(R)\left[c_{LR}\log_2\left(\frac{1}{c_{LR}}\right)+c_{HR}\log_2\left(\frac{1}{c_{HR}}\right)\right] \tag{6–51}$$

$$-P(F)\left[c_{LF}\log_2\left(\frac{1}{c_{LF}}\right)+c_{HF}\log_2\left(\frac{1}{c_{HF}}\right)\right] \tag{6–52}$$

and replacing the values of the transition matrix

$$=P(L)\log_2\left(\frac{1}{P(L)}\right)+P(H)\log_2\left(\frac{1}{P(H)}\right) \tag{6–53}$$

$$-P(R)\left[1\log_2 1+0\log_2 0\right] \tag{6–54}$$

$$-P(F)\left[\frac{1}{2}\log_2 2+\frac{1}{2}\log_2 2\right] \tag{6–55}$$

which simplifies to

$$=P(L)\log_2\left(\frac{1}{P(L)}\right)+P(H)\log_2\left(\frac{1}{P(H)}\right)-P(F) \tag{6–56}$$

replacing the values of $P(H), P(L)$, and $P(F)$ previously computed

$$=.6\log_2\left(\frac{1}{.6}\right)+.4\log_2\left(\frac{1}{.4}\right)-.8 \tag{6–57}$$

$$=.971-.8 \tag{6–58}$$

$$=.171 bits. \tag{6–59}$$

Solution to Problem 3, part f.

The maximum mutual information over all the possible input distributions gives us a quantity that depends only on the channel (as it should be if we are interested in computing the channel capacity); it tells us the maximum amount of information that could be transmitted through this channel (A good way to understand this is looking at the information flow diagrams introduced in chapter 7).

We can obtain M_{max} either graphically or differentiating M with respect to th einput probability distribution. Recall the last step in our derivation of the mutual information:

$$M=P(L)\log_2\left(\frac{1}{P(L)}\right)+P(H)\log_2\left(\frac{1}{P(H)}\right)-P(F) \tag{6–60}$$

We can express $P(H), P(L)$ in terms of $P(F)$, we already did so in part b. Let us rename $p=P(F)$, and express equation (6–60) in terms of p. After some manipulations we will reach the following expression

$$M=\log_2\left(\frac{2}{2-p}\right)+\frac{p}{2}\log_2\left(\frac{2-p}{p}\right)-p. \tag{6–61}$$

We just have to take a derivative of M with respect to p and equal the result to zero. Doing so will yield the following equation:

$$\frac{dM}{dp} = \frac{1}{2}\log_2\left(\frac{2-p}{p}\right) - 1 = 0 \tag{6–62}$$

After isolating p you should obtain that the maximum of M occurs at $p = P(F) = .4$. Note that after introducing asymmetric noise, the maximum mutual information is no longer centered at 0.5 as it was for the binary noiseless channel. Figure 6–3 illustrates this effect.

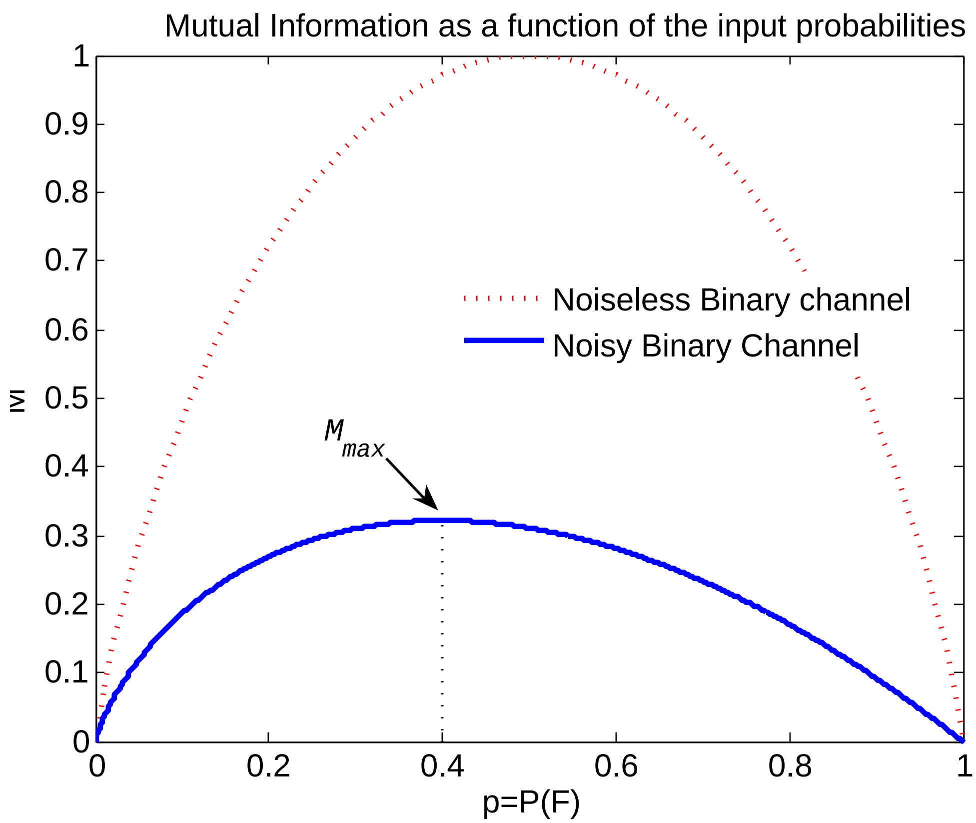

Figure 6–3: Comparison of the mutual information for noise-free and noisy channels

This is due to a competing effect between maximizing the amount of information transmitted and minimizing the errors introduced by the channel: since errors are no longer symmetric, minimizing the errors amounts to breaking the balance between symbols, but on the other hand breaking this balance tends to decrease the entropy before transmission. This tradeoff is what makes $M(p)$ have a maximum.

From the graph, or, replacing $p = .4$ in equation (6–61), we get $M_{max} = .3219$.

Solution to Problem 3, part g.

To compute the channel capacity now, we just need to plug the value of M_{max} in the formula for the capacity of a noisy channel given in the notes. $C = W \times M_{max} = .3219 bits/second$. So the manager has to be concerned. The noisy channel can only transmit .3219 bits/second of information, and information gets lost in the way. Note that if the system had been error-free there would have been no problem in keeping up with incoming customers.