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

6.050J/2.110J **Information and Entropy** **Spring 2006**

Issued: March 14, 2006

Problem Set 6 Solutions

Due: March 17, 2006

Solution to Problem 1: The Registrar's Worst Nightmare

Table 6–1 summarizes our knowledge at the beginning of the problem.

Probability	Value	Corresponding Statement
$p(U \mid A)$	0.15	15% of the Add Forms got lost in the mail
$p(E \mid A)$	0.85	[85% of the Add Forms were received properly]
$p(U \mid R)$	0.05	5% of the Registration Forms were illegible because of ink smears
$p(E \mid R)$	0.95	[95% of the Registration Forms were legible]
$p(E \mid N)$	0.01	errors in typing enrolled 1% of the people who didn't want to take the course
$p(U \mid N)$	0.88	[88% of the people who didn't want to take the course were not accidentally enrolled]

Table 6–1: Our Knowledge of the Probabilities

Solution to Problem 1, part a.

Reading from Table 6–1, we know that the probability that Alice is enrolled given that she submitted an Add form is $p(E \mid A) = 0.85$.

Solution to Problem 1, part b.

This part gives us the information that the probability a student submitted an Add form given that the student is enrolled is 20%. In probability-speak, this is $p(A \mid E) = 0.2$, the answer to the question.

Solution to Problem 1, part c.

Here we are given the information that $p(E) = 0.02$. From Bayes' theorem we know that $p(A, E) = p(A)p(E \mid A) = p(E)p(A \mid E)$. Since we have the last two numbers, we can calculate our answer as $p(A, E) = 0.02 \times 0.2 = 0.004$, or 0.4%.

Solution to Problem 1, part d.

From part b. we know that $p(A \mid E) = 0.2$. Thus, from part c. we have the equation:

$$p(A) = \frac{p(E)p(A \mid E)}{p(E \mid A)} \tag{6–1}$$

which is equal to $\frac{0.02 \times 0.2}{0.85} = 0.00047$, or 0.47%.

Solution to Problem 1, part e.

As before, $p(A \mid E) = 0.2$, or 20%.

Solution to Problem 1, part f.

We are looking for $p(N)$. We can write the following equations:

$$p(E) = p(E \mid A)p(A) + p(E \mid R)p(R) + p(E \mid N)p(N) \tag{6–2}$$

$$1 = p(A) + p(R) + p(N) \tag{6–3}$$

We can substitute Equation 6–3 into Equation 6–2 to eliminate $p(R)$ and obtain:

$$\begin{aligned} p(N) &= \frac{p(E) - p(E \mid A)p(A) - p(E \mid R) + p(E \mid R)p(A)}{p(E \mid N) - p(E \mid R)} \\ &= \frac{0.02 - 0.85 \times 0.0047 - 0.95 + 0.95 \times 0.0047}{0.02 - 0.95} \\ &= \frac{0.9295}{0.94} \\ &= 0.989 \end{aligned} \tag{6–4}$$

So the answer is 98.9%.

Solution to Problem 1, part g.

Here the probability is $p(E) = 0.02$, or 2%.

Solution to Problem 1, part h.

Here we are looking for $p(E|N)$, which we've known from the start to be 1%.

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."

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{6–5}$$

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

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

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

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.

$$H(loss) = -0.01 * \log_2(0.01) - 0.99 * \log_2(0.99) \quad (6\text{–}6)$$
$$= 0.0664 + 0.0143 \quad (6\text{–}7)$$
$$H(loss) = 0.0807 \quad (6\text{–}8)$$
$$(6\text{–}9)$$

$H(loss)$ 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.

$$H(transmitted) = 1 - 0.0807 = 0.9192 \quad (6\text{–}10)$$
$$W = 24000 bits/sec \quad (6\text{–}11)$$
$$C = 24000 * 0.9192 \quad (6\text{–}12)$$
$$= 22610 \text{ bits/sec.} \quad (6\text{–}13)$$

Solution to Problem 2, part b.

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

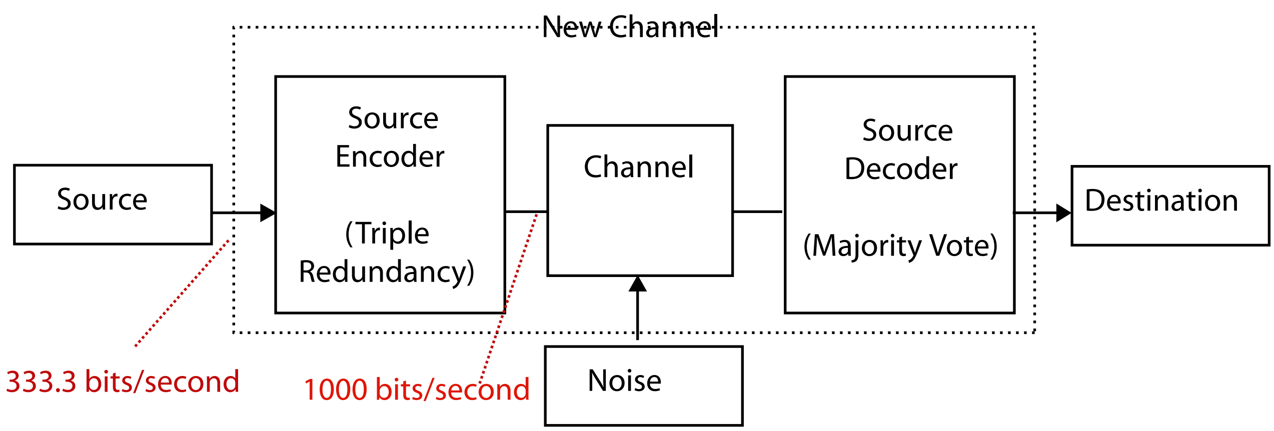

Figure 6–1: New channel model after addition of Hamming code

Since the original channel has a bandwidth of 24000 bit/sec and we are now using Hamming code [7,4], our new maximum channel capacity that we can feed into our new channel is $24000 \times 4/7 = 13714$ bits/sec.

Solution to Problem 2, part c.

Also, our channel characteristics change. No longer is the probability of a bit error 1%. Using Bayes theorem and the binomial distribution to establish the probability 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.

$$P(A, 0 \text{ errors}) = P(A|0 \text{ errors}) \times P(0 \text{ errors}) = 0 \quad (6\text{–}14)$$
$$P(A, 1 \text{ error}) = P(A|1 \text{ error}) \times P(1 \text{ error}) = 0 \quad (6\text{–}15)$$
$$P(A, 2 \text{ errors}) = P(A|2 \text{ errors}) \times P(2 \text{ errors}) \quad (6\text{–}16)$$
$$= 1 \times \binom{7}{2} \varepsilon^2 (1-\varepsilon)^5 \quad (6\text{–}17)$$
$$= 1 \times 21 * (0.01)^2 * (1 - 0.01)^5) = 0.0020 \quad (6\text{–}18)$$
$$P(A, 3 \text{ errors}) = P(A|3 \text{ errors}) * P(3 \text{ errors}) = \quad (6\text{–}19)$$
$$= 1 \times \binom{7}{3} \varepsilon^3 (1-\varepsilon)^4 \quad (6\text{–}20)$$
$$= 1 \times 35(0.01)^3 (1 - 0.01)^4) = 3.36 \times 10^{-5} \quad (6\text{–}21)$$
$$\vdots \quad (6\text{–}22)$$
$$P(A) = \sum_{i=0}^{7} P(A, i \text{ errors}) \quad (6\text{–}23)$$

Aparently, all the terms $P(A, i \text{ errors})$ for $i > 2$ are much smaller than $P(A, 2 \text{ errors})$ and we can ignore them in the sum $P(A)$; thus, giving an approximate value of $P(A) \approx 0.002$.

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.

$$H(loss) = 0.208 \quad (6\text{–}24)$$
$$(6\text{–}25)$$
$$H(transmitted) = 1 - 0.0208 = 0.9792 \quad (6\text{–}26)$$
$$W = 13714 bits/sec \quad (6\text{–}27)$$
$$C = 13714 * 0.9792 = 13429 \text{ bits/sec} \quad (6\text{–}28)$$

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.

If we repeat the same steps for the triple redundancy code, we will have:

$$P(A, 0 \text{ errors}) = P(A|0 \text{ errors}) \times P(0 \text{ errors}) = 0 \quad (6\text{–}29)$$
$$P(A, 1 \text{ error}) = P(A|1 \text{ error}) \times P(1 \text{ error}) = 0 \quad (6\text{–}30)$$
$$P(A, 2 \text{ errors}) = P(A|2 \text{ errors}) \times P(2 \text{ errors}) \quad (6\text{–}31)$$
$$= 1 \times \binom{3}{2} \varepsilon^2 (1-\varepsilon)^1 \quad (6\text{–}32)$$
$$= 1 \times 3 * (0.01)^2 * (1 - 0.01)) \quad (6\text{–}33)$$
$$P(A, 3 \text{ errors}) = P(A|3 \text{ errors}) * P(3 \text{ errors}) = \quad (6\text{–}34)$$
$$= 1 \times \binom{3}{3} \varepsilon^3 \quad (6\text{–}35)$$
$$= 1 \times (0.01)^3 \quad (6\text{–}36)$$
$$P(A) = P(A, 0 \text{ errors}) + P(A, 1 \text{ error}) + P(A, 2 \text{ errors}) + P(A, 3 \text{ errors}) \quad (6\text{–}37)$$
$$P(A) = 2.97 \times 10^{-4} + 10^{-6} \quad (6\text{–}38)$$
$$P(A) = 2.98 \times 10^{-4} \quad (6\text{–}39)$$

$$H(loss) = -2.98 \times 10^{-6} * \log_2(2.98 \times 10^{-6}) - (1 - 2.98 \times 10^{-6}) * \log_2((1 - 2.98 \times 10^{-6}) \quad (6\text{–}40)$$
$$H(loss) = 0.0039 \quad (6\text{–}41)$$
$$(6\text{–}42)$$
$$H(transmitted) = 1 - 0.0039 = 0.9961 \quad (6\text{–}43)$$
$$W = 24000 \times 1/3 = 8000 bits/sec \quad (6\text{–}44)$$
$$C = 8000 \times 0.9961 = 7968.8 bits/sec \quad (6\text{–}45)$$

Our probability of bit error in the Hamming code system is above the 0.1% 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 that reduces by the factor 4/7. The triple redundancy code has a lower rate of 1/3 but supresses the error well beyond the 0.1 % limit. The fundamental tradeoff between transmission rate and transmission quality is always encountetes in the communications theory.

Solution to Problem 3: Dealing with bad detectors

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

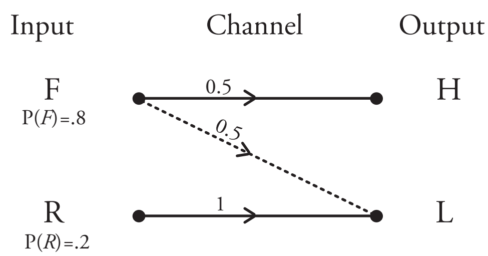

Figure 6–2: Alyssa's fiber optical communication system, transition probabilities are expressed on top of the lines connecting input with output.

The conclusion is that we are dealing with an instance of an asymmetric binary communication channel. We can 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:

$$\begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0 & 0.5 \end{bmatrix} \tag{6–46}$$

Solution to Problem 3, part b.

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

$$P(B_0) = P(0|0)P(A_0) + P(0|1)P(A_1) \quad = c_{00}P(A_0) + c_{01}P(A_1) = 1 \times .8 + 0.5 \times .2P(B_0) \quad = .9 \tag{6–47}$$

$$P(B_1) = P(1|0)P(A_0) + P(1|1)P(A_1) \quad = c_{10}P(A_0) + c_{11}P(A_1) = 0 \times .8 + 0.5 \times .2P(B_1) \quad = .1 \tag{6–48}$$

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: $P(1) = .2 = 1 - P(0)$. Then according to the lecture notes the uncertainty before transmission is (also known as entropy and often referred to with the letter H):

$$U_{\text{before}} = P(A_1)\log_. 2\left(\frac{1}{P(A_1)}\right) + P(A_0)\log_2\left(\frac{1}{P(A_0)}\right) \tag{6–49}$$

$$= .4644 + .2575 \tag{6–50}$$

$$U_{\text{before}} = .7219 \tag{6–51}$$

Solution to Problem 3, part d.

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

$$C = W \log_2 n \tag{6–52}$$

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:

$$C = 1 bit/second$$

.

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.

$$M = \sum_j P(B_j) \log_2 \left(\frac{1}{B_j} \right) - \sum_i P(A_i) \sum_j P(B_j|A_i) \log_2 \left(\frac{1}{P(B_j|A_i)} \right) \qquad (6\text{–}53)$$

adapting it to our problem

$$(6\text{–}54)$$

and replacing the values of the transition matrix

$$= P(B_0) \log_2 \left(\frac{1}{P(B_0)} \right) + P(B_1) \log_2 \left(\frac{1}{P(B_1)} \right) \qquad (6\text{–}55)$$

$$- P(A_0) \left[1 \log_2 1 + 0 \log_2 0 \right] \qquad (6\text{–}56)$$

$$- P(A_1) \left[\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 \right] \qquad (6\text{–}57)$$

which simplifies to

$$= P(B_0) \log_2 \left(\frac{1}{P(B_0)} \right) + P(B_1) \log_2 \left(\frac{1}{P(B_1)} \right) - P(A_1) \qquad (6\text{–}58)$$

replacing the values of $P(B_0), P(B_1)$, and $P(A_1)$ previously computed

$$= .9 \log_2 \left(\frac{1}{.9} \right) + .1 \log_2 \left(\frac{1}{.1} \right) - .2 \qquad (6\text{–}59)$$

$$= .4690 - .2 \qquad (6\text{–}60)$$

$$= .2690 \text{ bits.} \qquad (6\text{–}61)$$

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:

$$M = P(B_0)\log_2\left(\frac{1}{P(B_0)}\right) + P(B_1)\log_2\left(\frac{1}{P(B_1)}\right) - P(A_1) \tag{6–62}$$

We can express $P(B_0), P(B_1)$ in terms of $P(A_1)$, we already did so in part b. Let us rename $p = P(A_1)$, and express equation (6–62) in terms of p. After some manipulations we will reach the following expression

$$M = \log_2\left(\frac{2}{2-p}\right) + \frac{p}{2}\log_2\left(\frac{2-p}{p}\right) - p. \tag{6–63}$$

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:

$$\frac{dM}{dp} = \frac{1}{2}\log_2\left(\frac{2-p}{p}\right) - 1 = 0 \tag{6–64}$$

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.

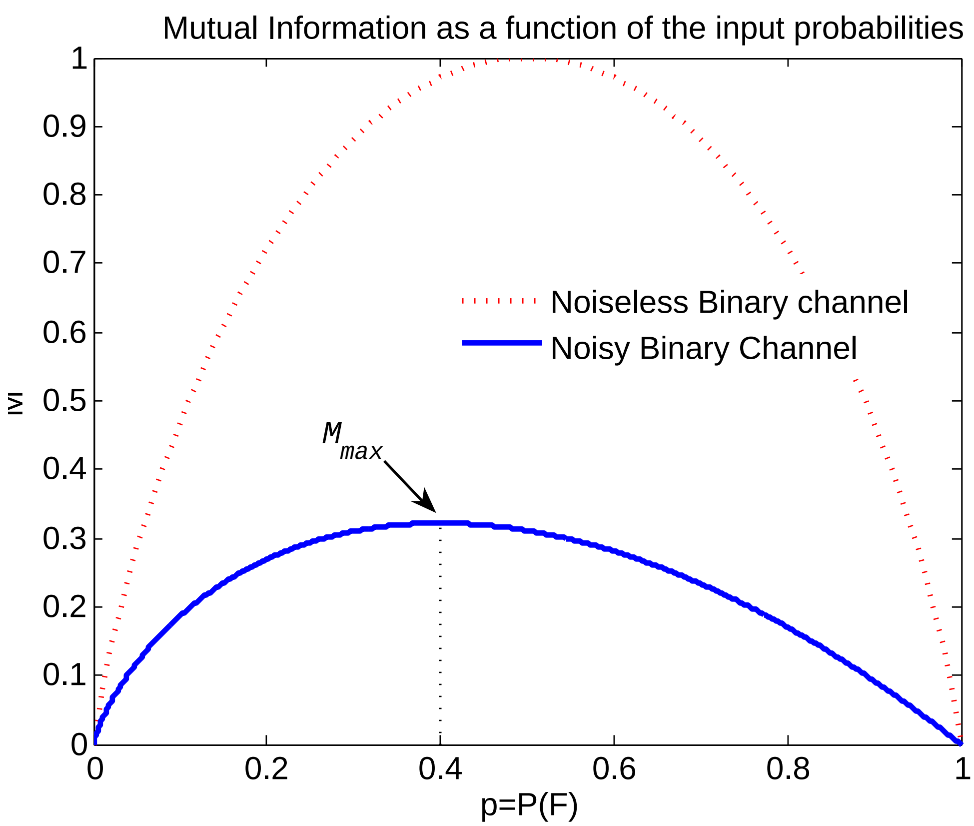

Figure 6–3: Comparison of the mutual information for noise-free and noisy channels

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 (6–63), we get $M_{max} = .3219$. So Alyssa has to be more carefull about her choice of source coder as with the choice of $p = 0.2$ that she has made, the system does not acheive the capacity of the channel.

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. $C = W \times M_{max} = 10^9 .3219 bits/second = 3.219 \times 10^8 = 321.9 Mbps$.