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

6.050J/2.110J **Information and Entropy** **Spring 2005**

Quiz Solutions

Solution to Problem 1: (25%)

In computing cumulative ratings (MIT's version of GPA) grades of A count as 5, grades of B as 4, C as 3, and so on. One semester Professor Plum gave only grades of A, B, and C in his course "Process of Elimination." Later he remembered that the average grade he assigned, using the numerical scale above, was 4.30 but he did not have a clue about the actual number of A, B, or C grades. Without making any further assumptions, you want to estimate the fraction of A, B, and C grades (i.e., the probabilities that a student selected at random would have received a particular grade). For simplicity you can refer to these unknown probabilities as A, B, and C, rather than $p(A)$, $p(B)$, and $p(C)$.

Solution to Problem 1, part a.

To find the possible ranges of A, B, and C, we first to express them in terms of other variables, using the constraints that: $A+B+C=1$, and $5\times A+4\times B+3\times C=4.3$ given in the problem statement, we obtain the following set of equations:

$$\begin{array}{llll} A & = .3+C & A & = .65-\frac{B}{2} \\ B & = .7-2C & B & = 1.3-2A \\ C & = A-.3 & B & = .35-\frac{B}{2} \end{array} \tag{Q–1}$$

With these equations we can establish the valid ranges for the three variables

$$\underset{A_{\text{Min}}}{\underline{\quad .3 \quad}} \le A \le \underset{A_{\text{Max}}}{\underline{\quad .65 \quad}} \qquad \underset{B_{\text{Min}}}{\underline{\quad 0 \quad}} \le B \le \underset{B_{\text{Max}}}{\underline{\quad .7 \quad}} \qquad \underset{C_{\text{Min}}}{\underline{\quad 0 \quad}} \le C \le \underset{C_{\text{Max}}}{\underline{\quad .35 \quad}}$$

Solution to Problem 1, part b.

Next, we asked you to determine the expression of the entropy that would have to be maximized to obtain the distribution of grades, in terms of only one of the probabilities. This is easily done using the relations given in equation Q–1. You had the choice to express the entropy in terms of either A, B or C. All the possible answers are given below

$$\text{Entropy} = \underline{(.3+C)\log\left(\frac{1}{.3+C}\right) + (.7-2C)\log\left(\frac{1}{.7-2C}\right) + C\log\left(\frac{1}{C}\right)}$$

$$\text{Entropy} = \underline{(.65+.5B)\log\left(\frac{1}{.65+.5B}\right) + B\log\left(\frac{1}{B}\right) + (.35-.5B)\log\left(\frac{1}{.35-.5B}\right)}$$

$$\text{Entropy} = \underline{A\log\left(\frac{1}{A}\right) + (1.3-2A)\log\left(\frac{1}{1.3-2A}\right) + (A-.3)\log\left(\frac{1}{A-.3}\right)}$$

Solution to Problem 2: (20%)

According to the Registrar, Mrs. White, the grades that semester were actually 50% A, 30% B, and 20% C. You decide to produce an encoding to store the thousands of grades assigned by Professor Plum efficiently. You want to use a Huffman code for this purpose, because you can represent the grades using fewer bits on average than two per grade, which would be required for a fixed-length code.

Solution to Problem 2, part a.

The entropy per assigned grade of this source of data, is simply the entropy of the distribution of grades. That is

$$S = -.5 \log .5 - .3 \log .3 - .2 \log .2. \tag{Q–2}$$

Hence, the correct solution for this part was:

0.92 bits   1.12 bits   1.32 bits $\boxed{\text{1.49 bits}}$ 1.74 bits   2.32 bits

Solution to Problem 2, part b.

Coming up with a Huffman Code for just three symbols is simpler than doing so for large symbol space sizes. The two symbols with smaller probability (B and C in this case) get assigned a two bit code (making sure that the prefix-free condition holds, that is that no two codes have the same prefix), the symbol with higher probability (A in this case) is assigned a one bit code (still inforcing the prefix-free condition). The resulting average codeword length is the expected value of these lengths given the input probability distribution:

$$L_{av} = A \times 1\ bit + B \times 2\ bit + C \times 2\ bit \tag{Q–3}$$
$$= .5 \times 1 + .3 \times 2 + .2 \times 2 \tag{Q–4}$$
$$= .5 + .6 + .4 \tag{Q–5}$$
$$= 1.5\ bits \tag{Q–6}$$

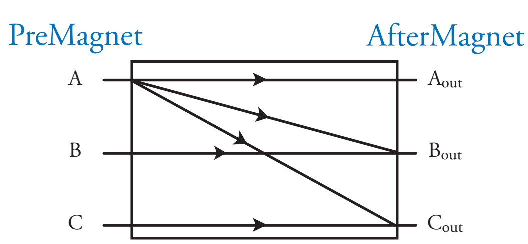
Based on this one possible Huffman code is

A __1__   B __01__   C __00__   Average Length __1.5 bits__

Note how the average length thus obtained is within one bit of the input entropy you computed in previous parts.

Solution to Problem 3: (20%)

Professor Plum's teaching assistant, Miss Scarlet, accidentally damaged a floppy disk containing the grades (she did it in the lab with a magnet). Many random errors were introduced – 40% the grades of A were shifted to B, and 20% of the grades of A became C. Grades of B and C were not affected.



The police investigator, Colonel Mustard, treated this incident as a communication channel with errors. Using the probability distribution from Problem 2 ($p(A) = 0.5$, $p(B) = 0.3$, and $p(C) = 0.2$), he calculated the probability distribution of the grades on the damaged disk A_{out}, B_{out}, and C_{out} and the resulting entropy at the output J. He reasoned that if the output information is the same as the input information there was nothing lost and no need to make any corrections, so the case could be closed. What did he find? Do you agree with his reasoning?

In this case we have to compute the output probabilities given the channel. First we compute the transition probability matrix based on the figure and the numbers given in the paragraph:

$$c_{ji} = P(B_j|A_i) = \begin{bmatrix} .4 & 0 & 0 \\ .4 & 1 & 0 \\ .2 & 0 & 1 \end{bmatrix} \tag{Q–7}$$

To compute the output probabilities we need to use Bayes rule to obtain the joint distribution and then sum over all possible input values, that is:

$$P(B_j) = \sum_i P(B_j, A_i) = \sum_i P(B_j|A_j)P(A_j) \tag{Q–8}$$

Using this formula with the transition probability matrix derived above, yields the following solution:

$p(A_{out})$ __.2__ $p(B_{out})$ __.5__ $p(C_{out})$ __.3__ J __1.49 bits__

Note that since the distribution at the output assigns the same probabilities to different variables (that is, .5, .3, and .2), we do not need to recompute the entropy, J will be the same as I, 1.49 bits.

Concerning the reasoning of Colonel Mustard, the answer is:

Do you agree? __NO__ Why?: Knowledge of the mutual information is required to make such a statement. The process had both noise and loss, therefore the mutual information will differ from I.

Solution to Problem 4: (34%)

Solution to Problem 4, part a.

One of the students, Mr. Green, was told that this defective floppy disk had his grade as B. He asked you to estimate the probability that his originally assigned grade had been A, B, or C. You should make use of the original probability distribution, the error probabilities caused by the accident, and the fact that his grade is now shown as a B.

In this part we are asked to use Bayesian inference to estimate the probability of each input given that the output was B_{out}. That is for each input A_i we have to compute

$$P(A_i|B_{out}) = \frac{P(B_{out}|A_i)P(A_i)}{P(B_{out})} \tag{Q–9}$$

In this case, the refined input probabilities result from applying the above formula with A_i as A, B, and C, :

$$P(A|B_{out}) = \frac{P(B_{out}|A)P(A)}{P(B_{out})} \quad \text{(Q–10)}$$
$$= \frac{.4 \times .5}{.5} \quad \text{(Q–11)}$$
$$= .4 \quad \text{(Q–12)}$$
$$P(B|B_{out}) = \frac{P(B_{out}|B)P(B)}{P(B_{out})} \quad \text{(Q–13)}$$
$$= \frac{\times .3}{.5} \quad \text{(Q–14)}$$
$$= .6 \quad \text{(Q–15)}$$
$$P(C|B_{out}) = \frac{P(B_{out}|C)P(C)}{P(B_{out})} \quad \text{(Q–16)}$$
$$= \frac{0 \times .2}{.5} \quad \text{(Q–17)}$$
$$= 0 \quad \text{(Q–18)}$$

Abusing a little bit of notation we identify these three probabilities with the new refined input probabilities

$p(A)$ 0.4 $p(B)$ 0.6 $p(C)$ 0

Solution to Problem 4, part b.

You decide to review Colonel Mustard's conclusions. Calculate the residual uncertainty U about the original grade for each of the three possible output grades, $U_{before}(A_{out})$, $U_{before}(B_{out})$, and $U_{before}(C_{out})$, and from those the loss L and mutual information M.

To compute the entropy at the input given the outputs,

$$U_{before}(B_j) = \sum_i P(A_i|B_j) \log\left(\frac{1}{P(A_i|B_j)}\right), \quad \text{(Q–19)}$$

we have to compute $P(A_i|A_{out})$ as we did in the last part for $P(A_i|B_{out})$.

$$P(A|A_{out}) = \frac{P(A_{out}|A)P(A)}{P(A_{out})} \quad \text{(Q–20)}$$
$$= \frac{.4 \times .5}{.2} \quad \text{(Q–21)}$$
$$= 1 \quad \text{(Q–22)}$$
$$P(B|A_{out}) = \frac{P(A_{out}|B)P(B)}{P(A_{out})} \quad \text{(Q–23)}$$
$$= \frac{0 \times .3}{.2} \quad \text{(Q–24)}$$
$$= 0 \quad \text{(Q–25)}$$
$$P(C|A_{out}) = \frac{P(A_{out}|C)P(C)}{P(A_{out})} \quad \text{(Q–26)}$$
$$= \frac{0 \times .2}{.2} \quad \text{(Q–27)}$$
$$= 0 \quad \text{(Q–28)}$$
$$\quad \text{(Q–29)}$$

and also, $P(A_i|C_{out})$

$$P(A|C_{out}) = \frac{P(C_{out}|A)P(A)}{P(C_{out})} \tag{Q–30}$$
$$= \frac{.2 \times .5}{.3} \tag{Q–31}$$
$$= \frac{1}{3} \tag{Q–32}$$
$$P(B|C_{out}) = \frac{P(C_{out}|B)P(B)}{P(C_{out})} \tag{Q–33}$$
$$= \frac{0 \times .3}{.3} \tag{Q–34}$$
$$= 0 \tag{Q–35}$$
$$P(C|C_{out}) = \frac{P(C_{out}|C)P(C)}{P(C_{out})} \tag{Q–36}$$
$$= \frac{1 \times .2}{.3} \tag{Q–37}$$
$$= \frac{2}{3} \tag{Q–38}$$
$$\tag{Q–39}$$

Quick inspection of $P(\cdot|A_{out})$ shows us that $U_{before}(A_{out}) = 0$. The other two conditional entropies at the input can be readily computed using the logarithm table provided with the exam. The resulting entropies are

$U_{before}(A_{out})$ __0 bits__ $U_{before}(B_{out})$ __.97 bits__ $U_{before}(C_{out})$ __.92 bits__

The Loss is defined as the average entropy at the input given the outputs:

$$L = P(A_out) \times U_{before}(A_{out}) + P(B_{out}) \times U_{before}(B_{out}) + P(C_{out}) \times U_{before}(C_{out}) \tag{Q–40}$$

Everything we need to compute the loss has already been computed, so simple substitution in this equation yields:

$$L = .2 \times 0 + .5 \times 0.97 + .3 \times .92 = .761 \tag{Q–41}$$

We can now compute the mutual information making use of the equality $M = I - L = 1.49 - 0.761 = 0.729$. The same result can be achieved computing noise and comparing with the outpuit information. In fact, since input and output information are the same, we know that Loss and Noise must be the same as well. Your solutions should then have looked like

L __0.761__ M __0.729__