PROBLEM 1. Since the class of instantaneous codewords is a subset of the class of uniquely decodable codewords, it follows that $\bar{M}_2 \leq \bar{M}_1$. On the other hand, let $\{l_i\}$ be the codeword lengths of the uniquely decodable code for which $\bar{M} = \bar{M}_2$. Since $\{l_i\}$ satisfies the Kraft's inequality, there exists an instantaneous code with these codeword lengths. For this instantaneous code $\bar{M} = M_2$ and we see that $\bar{M}_1 \leq \bar{M} = M_2$, and we conclude that $\bar{M}_1 = \bar{M}_2$.

PROBLEM 2. (a) 1/3

(b) The probability is 2/3. One easy way to see this is the following. If the easy question was picked initially, then it is in the student's box and this happens with probability 1/3. Otherwise, if a very hard question was picked initially, then the easy question is in the remaining box, and this happens with probability 2/3. The reason why the probabilities change is that revealing a box with a very hard question gives the student some information about the other boxes: the choice of the box being opened is not independent from the content of the box.

PROBLEM 3.

(a) $\{00, 01, 100, 101, 1100, 1101, 1110, 1111\}$.

(b) First note that if any two number differ by 2^{-k}, their binary expansion will differ somewhere in the first k bits after the 'point'. (Think of the decimal case: if $a = 0.375\ldots$ and b differs by more than 10^{-3} by it, then b's expansion cannot start with 0.375.)

Next observe that that for $i > j$

$$Q_i - Q_j = \sum_{k=j}^{i-1} P(a_k) \geq P(a_j) \geq 2^{-l_j}.$$

So, the binary expansion of Q_i and Q_j must differ somewhere in the first l_j bits. Since codewords for i and j are at least l_j bits long, neither codeword can be a prefix of the other.

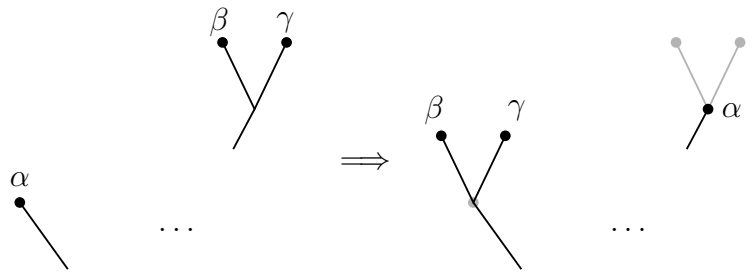
The bound on the average codeword length follows from

$$-\log_2 P(a_i) \leq l_i < -\log_2 P(a_i) + 1.$$

This method of coding is also known as Shannon coding and predates Huffman coding.

PROBLEM 4.

(a) Consider the longest and the shortest codewords. We know that there are at least two longest codewords, suppose their length is l. Suppose the shortest codewords has length s. Suppose that s and l differ by 2 or more. To show that this cannot be the case for an optimal code, consider the transformation shown below:



We see that the transformation decreases the length of two codewords (for letters β and γ) by $l-(s+1)=l-s-1$, whereas it increases the length of one codeword (for the letter α) by $(l-1)-s=l-s-1$. But since $l-s-1>0$, and since all the codewords are equally likely, this would have decreased the average codeword length, contradicting the optimality of the Huffman code. Thus, the longest and shortest codeword lengths can differ by at most 1, and these lengths must be j and $j+1$. (If some other two consecutive depths were used we would either not have enough leaves, or have too many leaves).

(b) Let the number of codewords of length k be m_k, $k=j,j+1$. Since the Huffman procedure yields a complete tree (no leaf is unoccupied) all intermediate nodes have two children. Thus, the 2^j nodes at level j of the tree are either codewords (m_j of them) or each of their two children are codewords ($m_{j+1}/2$ of them). Thus

$$m_j + m_{j+1}/2 = 2^j,$$

and also $m_j + m_{j+1} = x2^j$. From these two equations we find

$$m_j = (2-x)2^j \quad \text{and} \quad m_{j+1} = (x-1)2^{j+1}.$$

(c) By the result of (b) the average codeword length is

$$[jm_j + (j+1)m_{j+1}]/(x2^j) = j + 2(x-1)/x.$$

Problem 5.

(a) $H(X) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = 0.918$ bits $= H(Y)$.

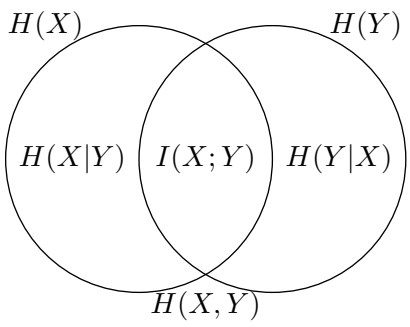
(b) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = 0.667$ bits $= H(Y|X)$.

(c) $H(X,Y) = 3 \times \frac{1}{3}\log 3 = 1.585$ bits.

(d) $H(Y) - H(Y|X) = 0.251$ bits.

(d) $I(X;Y) = H(Y) - H(Y|X) = 0.251$ bits.

(f)



Problem 6.

$$\begin{aligned} H(X) &= -\sum_{k=1}^{M} P_X(a_k) \log P_X(a_k) \\ &= -\sum_{k=1}^{M-1} (1-\alpha) P_Y(a_k) \log[(1-\alpha) P_Y(a_k)] - \alpha \log \alpha \\ &= (1-\alpha) H(Y) - (1-\alpha) \log(1-\alpha) - \alpha \log \alpha \end{aligned}$$

Since Y is a random variable that takes $M-1$ values $H(Y) \leq \log(M-1)$ with equality if and only if Y takes each of its possible values with equal probability.

Problem 7.

(a) Using the chain rule for mutual information,

$$I(X,Y;Z) = I(X;Z) + I(Y;Z \mid X) \geq I(X;Z),$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

(b) Using the chain rule for conditional entropy,

$$H(X,Y \mid Z) = H(X \mid Z) + H(Y \mid X,Z) \geq H(X \mid Z),$$

with equality iff $H(Y \mid X,Z) = 0$, that is, when Y is a function of X and Z.

(c) Using first the chain rule for entropy and then the definition of conditional mutual information,

$$\begin{aligned} H(X,Y,Z) - H(X,Y) = H(Z \mid X,Y) &= H(Z \mid X) - I(Y;Z \mid X) \\ &\leq H(Z \mid X) = H(X,Z) - H(X)\,, \end{aligned}$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

(d) Using the chain rule for mutual information,

$$I(X;Z \mid Y) + I(Z;Y) = I(X,Y;Z) = I(Z;Y \mid X) + I(X;Z)\,,$$

and therefore

$$I(X;Z \mid Y) = I(Z;Y \mid X) - I(Z;Y) + I(X;Z)\,.$$

We see that this inequality is actually an equality in all cases.