SIMG-714 Information Theory for Imaging Science
Homework 2

1. An information source X has M symbols. Show that the source entropy is bounded by $H(X) \leq \log_2 M$.Explain the conditions under which equality holds.

   Consider the difference

   $$H(X) - \log M = \sum_{k=1}^{M} p_k \log\left(\frac{1}{Mp_k}\right) \leq \sum_{k=1}^{M} p_k \left[\frac{1}{Mp_k} - 1\right] \ln 2 = \left(\sum_{k=1}^{M} \frac{1}{M} - \sum_{k=1}^{M} p_k\right) \ln e = 1 - 1 = 0$$

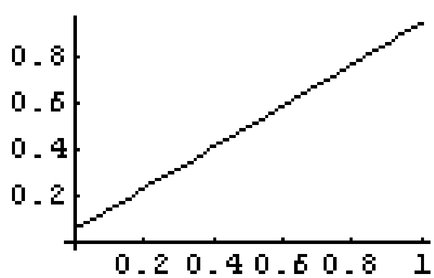
   Hence, $H(X) \leq \log M$ with equality if and only if $p_k = 1/M$ for $1 \leq k \leq M$.

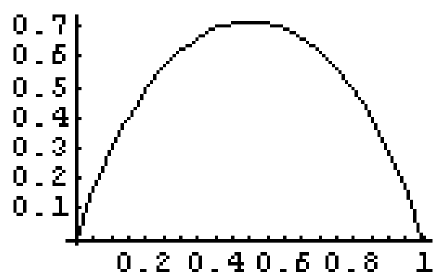
2. Compute the average information transmission through a binary symmetric communication channel with source probabilities $P(x_1) = P(x_2) = 0.5$ and channel error probability $P(y_2|x_1) = P(y_1|x_2) = 0.9$. Explain your result.

   The average information transmission per symbol can be computed by $I(X;Y) = H(Y) - H(Y|X)$. Compute $p(y_1) = (0.5)(0.1)+(0.5)(0.9) = 0.5$ and $p(y_2) = 1-p(y_2) = 0.5$ so that $H(Y) = 1$ bit/symbol. Using the channel transition probabilities find $H(Y|x_1) = H(Y|x_2) = -0.1 \log 0.1 - 0.9 \log 0.9 = 0.469$ bit/symbol. Then $H(Y|X) = (0.5)(0.469) + (0.5)(0.469) = 0.469$. The information transmission is $I(X;Y) = H(Y) - H(Y|X)1 - 0.469 = 0.531$ bits/symbol. Although the crossover probabilities are greater than 0.5, this does not damage the information transmission capability of the channel. The same rate as is produced with $q = 0.1$ is obtained. The receiver merely needs to change its interpretation of the output.

3. Plot the amount of information that gets through a BSC with crossover probability $q = 0.05$ in terms of the probability $P(x_1) = p$. Explain the symmetry of your result.

   By the same method as above, we find that $p(y_1) = p(1-q)+(1-p)q$ and $p(y_2) = (1-p)(1-q)+pq$ and $H(Y|X) = -(q \log q + (1-q)\log(1-q))$. At $q = 0.05$ we find $HY|X) = 0.286$.The information through the channel as a function of p is $I(X;Y) = H(Y) - 0.286$. The entropy is computed using the probabilities $p(y_1)$ and $p(y_2) = 1 - p(y_1)$ as $H(Y) = -p(y_1)\log_2 p(y_1) - p(y_2)\log_2 p(y_2)$. As p ranges over [0,1] we see that $p(y_1)$ ranges over [0.05,0.95]. Hence, $H(Y)$ is never smaller than $H(Y|X)$. The peak value occurs where $p(y_1) = 0.5$, where $H(Y) = 1$ and $I(X;Y) = 1 - 0.286 = 0.714$. The probability $p(y_1)$ is plotted below left and the information transfer $I(X;Y)$ is plotted below right.





4. A binary symmetric channel has erasure probability $q = 0.05$. Plot the amount of information that gets through the channel in terms of the probability $P(x_1) = p$.

   The transition probabilities for this channel are listed in the $\mathcal{P}$ matrix below.

   $$\mathcal{P} = \begin{bmatrix} 1-q & q & 0 \\ 0 & q & 1-q \end{bmatrix}$$

   The joint probabilities $P(x,y)$ are then found by multiplying the first row by p and the second by $(1-p)$.

   $$\mathcal{P}_{XY} = \begin{bmatrix} p(1-q) & pq & 0 \\ 0 & (1-p)q & (1-p)(1-q) \end{bmatrix}$$

The probabilities $P(y)$ are found by summing the columns of the $\mathcal{P}_{XY}$ array.

$$P_Y = \left[\begin{array}{ccc} p(1-q) & q & (1-p)(1-q) \end{array}\right]$$

The average information transfer through the channel is $I(X;Y) = H(Y) - H(Y|X)$. The second term is

$$H(Y|X) = -\sum_{X,Y} p(x,y) \log p(y|x)$$

The elements in the expression are corresponding terms from the arrays $\mathcal{P}$ and $\mathcal{P}_{XY}$. The result is

$$\begin{aligned} H(Y|X) &= -p(1-q)\log(1-q) - pq\log q - (1-p)\log q - (1-p)(1-q)\log(1-q) \\ &= -q\log q - (1-q)\log(1-q) \end{aligned}$$

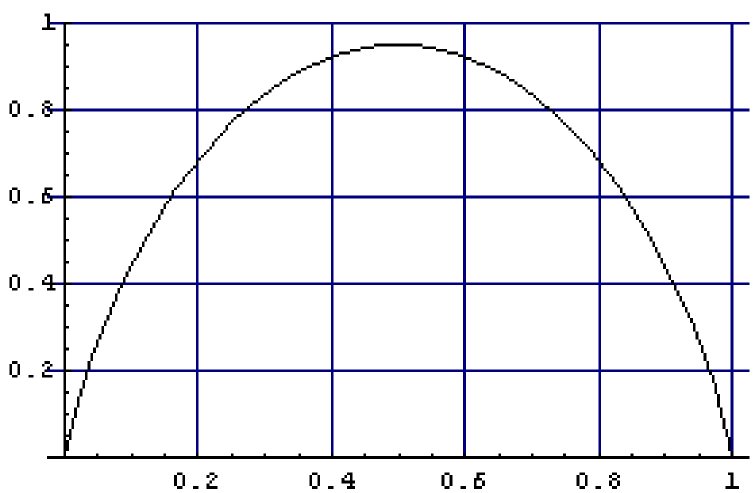
This turns out to be independent of the input probabilities because the channel is uniform from the input. The entropy $H(Y)$ can be computed as

$$\begin{aligned} H(Y) &= -\sum_Y p(y)\log p(y) \\ &= -p(1-q)\log p(1-q) - q\log q - (1-p)(1-q)\log(1-p)(1-q) \\ &= (1-q)\left[-p\log p - (1-p)\log(1-p)\right] - q\log q - (1-q)\log(1-q) \\ &= (1-q)\left[-p\log p - (1-p)\log(1-p)\right] + H(Y|X) \end{aligned}$$

Hence, the information transmission reduces to

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) = (1-q)\left[-p\log p - (1-p)\log(1-p)\right] \\ &= (1-q)H(X) \end{aligned}$$

The interpretation is that the erasures remove all information when they occur. When they do not occur, all of the information from the source gets through. The information transfer is clearly maximized by maximizing $H(X)$, which requires $p = 0.5$.



5. A discrete channel is characterized by the matrix

$$\mathcal{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$$

(a) Given the input probabilities $p(x_1) = 1/2$, $p(x_2) = p(x_3) = 1/4$, find $I(X;Y)$.
The joint probability array is found by multiplying each row by the appropriate $p(x_i)$.

$$\mathcal{P}_{XY} = \begin{bmatrix} \frac{1}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{24} & \frac{1}{8} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{24} & \frac{1}{8} \end{bmatrix}$$

The probabilities P_Y are found by summing the columns of $\mathcal{P}_{XY}$.

$$\mathcal{P}_Y = \left[\begin{array}{ccc} \frac{3}{8} & \frac{1}{3} & \frac{7}{24} \end{array}\right]$$

The average information transmission is given by $I(X;Y) = H(Y) - H(Y|X)$. Because the channel is uniform, $H(Y|X)$ can be computed from the probabilities in any row of $\mathcal{P}$.

$$H(Y|X) = \frac{1}{2}\log_2 2 + \frac{1}{3}\log_2 3 + \frac{1}{6}\log_2 6 = 1.459$$

$$H(Y) = -\frac{3}{8}\log_2\frac{3}{8} - \frac{1}{3}\log_2\frac{1}{3} - \frac{7}{24}\log_2\frac{7}{24} = 1.577$$

Hence, $I(X;Y) = 1.577 - 1.459 = 0.118$ bit per symbol.

(b) Construct the decision rules for an ideal observer and compute the error probability. An ideal observer chooses x_i to maximize $P(x_i|y_j)$ for each y_j.

The ideal observer computes the most probable input symbol for each output symbol that is observed. Hence, we need the matrix $\mathcal{P}_{X|Y}$. This can be found by dividing each element of $\mathcal{P}_{XY}$ by the value of $p(y_j)$ for that column. The result is the array

$$\mathcal{P}_{X|Y} = \left[\begin{array}{ccc} \frac{2}{3} & \frac{1}{2} & \frac{2}{7} \\ \frac{1}{9} & \frac{3}{8} & \frac{2}{7} \\ \frac{2}{9} & \frac{1}{8} & \frac{3}{7} \end{array}\right]$$

The ideal observer chooses the value of x according to the largest probability in each column. These turn out to be the elements on the diagonal. Hence, the decision rule is: $y_1 \to x_1$, $y_2 \to x_1$, $y_3 \to x_3$. The error probability can be computed as

$$\begin{aligned} P(E) &= P(y_1)P(E|y_1) + P(y_2)P(E|y_2) + P(y_3)P(E|y_3) \\ &= \left(\frac{3}{8}\right)\frac{1}{3} + \frac{1}{3}\frac{1}{2} + \frac{7}{24}\frac{4}{7} = \frac{11}{24} = 0.458 \end{aligned}$$

(c) Compute the channel capacity for this channel.

The channel is uniform from the input and output. Therefore, the capacity occurs when the input symbols are equally probable. This causes the output symbols to be equally probable as well. Under these conditions,

$$H(Y) = \log_2 3 = 1.585$$

The value of $H(Y|X)$ does not change. Hence,

$$C = 1.585 - 1.459 = 0.126 \text{ bits per symbol}$$

6. A number of identical binary symmetric channels are cascaded, so that the output of one channel is the input to the next. Let p_0 be the probability that $x = 0$ is the input to the first channel. Find the probability p_n that the input to channel n is 0. Find an explicit expression for p_n in terms of p_0 and the crossover probability, β, of a single BSC stage. Show that $p_n \to 1/2$ regardless of the value of p_0 if $\beta > 0$ when n becomes large.

This problem can be addressed in a number of ways. The most elegant, which also anticipates the next problem, is to use a matrix formulation. Let

$$\mathbf{P} = \left[\begin{array}{cc} 1-\beta & \beta \\ \beta & 1-\beta \end{array}\right]$$

and let $\boldsymbol{\pi}_0 = \left[\begin{array}{cc} p_0 & 1-p_0 \end{array}\right]$ be the input probabilities at the input to the first stage. Then the probabilities at the output of stage 1, which are the input to stage 2, are

$$\boldsymbol{\pi}_1 = \left[\begin{array}{cc} p_0 & 1-p_0 \end{array}\right]\left[\begin{array}{cc} 1-\beta & \beta \\ \beta & 1-\beta \end{array}\right] = \boldsymbol{\pi}_0\mathbf{P}$$

Then the output of stage 1 and input to stage 2 is $\boldsymbol{\pi}_2 = \boldsymbol{\pi}_1\mathbf{P} = \boldsymbol{\pi}_0\mathbf{P}^2$. In general, $\boldsymbol{\pi}_2 = \boldsymbol{\pi}_0\mathbf{P}^n$. The problem now is to compute $\mathbf{P}^n$.

The power of a matrix can be computed easily if it can be diagonalized. This is always possible if the matrix is real and symmetric, which is the case for $\mathbf{P}$. Let $\mathbf{Q}$ be a matrix such that

$$\mathbf{Q}^{-1}\mathbf{P}\mathbf{Q} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix. Then

$$\mathbf{P} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1}$$

By direct multiplication we find $\mathbf{P}^n = \left(\mathbf{QDQ}^{-1}\right)\left(\mathbf{QDQ}^{-1}\right)\cdots\left(\mathbf{QDQ}^{-1}\right)$. After regrouping we can write $\mathbf{P}^n = \mathbf{QD}\left(\mathbf{Q}^{-1}\mathbf{Q}\right)\mathbf{D}\left(\mathbf{Q}^{-1}\mathbf{Q}\right)\mathbf{D}\cdots\left(\mathbf{Q}^{-1}\mathbf{Q}\right)\mathbf{DQ}^{-1} = \mathbf{QD}^n\mathbf{Q}^{-1}$ because all of the internal $\mathbf{Q}^{-1}\mathbf{Q}$ terms are identies. It is a simple matter to compute $\mathbf{D}^n$ for a diagonal matrix.

$$\mathbf{D}^n = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}^n = \begin{bmatrix} d_1^n & 0 \\ 0 & d_2^n \end{bmatrix}$$

Then

$$\mathbf{P}^n = \mathbf{Q}\begin{bmatrix} d_1^n & 0 \\ 0 & d_2^n \end{bmatrix}\mathbf{Q}^{-1}$$

For a real symmetric matrix such as $\mathbf{P}$, the columns of $\mathbf{Q}$ are the normalized eigenvectors and the elements of $\mathbf{D}$ are the eigenvalues. The eigenvalues are the solution of the determinant

$$|\mathbf{P}-\lambda\mathbf{I}| = 0$$

$$\left|\begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right| = \begin{vmatrix} 1-\beta-\lambda & \beta \\ \beta & 1-\beta-\lambda \end{vmatrix} = 1 - 2\beta - 2\lambda + 2\beta\lambda + \lambda^2 = 0$$

The equation on the right can be factored

$$1 - 2\beta - 2\lambda + 2\beta\lambda + \lambda^2 = (\lambda - 1)(\lambda - 1 + 2\beta)$$

which yields the eigenvalues $\{\lambda_1 = 1, \lambda_2 = 1 - 2\beta\}$. The eigenvectors are found by solving the equation

$$\mathbf{P}\mathbf{x}_i = \lambda_i\mathbf{x}_i$$

$$\begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} = \lambda_i\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}$$

For the first eigenvalue,

$$\begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$$

This is equivalent to the simultaneous equations

$$\begin{aligned} (1-\beta)\,x_{11} + \beta x_{12} &= x_{11} \\ \beta x_{11} + (1-\beta)\,x_{12} &= x_{12} \end{aligned}$$

These equations have the solutions $[x_{11}, x_{12}] = [\alpha, \alpha]$ for any α. Choose $\alpha = 1/\sqrt{2}$ to normalize the first eigenvector, $\mathbf{x}_1 = \left[1/\sqrt{2}, 1/\sqrt{2}\right]^T$. The second equation is

$$\begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}\begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = (1-2\beta)\begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$$

This has the solution $[x_{21}, x_{22}] = [\alpha, -\alpha]$ for any α. Choose $\alpha = 1/\sqrt{2}$ to normalize the second eigenvector, $\mathbf{x}_2 = \left[1/\sqrt{2}, -1/\sqrt{2}\right]^T$.

The diagonalization matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Very conveniently, $\mathbf{Q}^{-1} = \mathbf{Q}$. The diagonal matrix is

$$\mathbf{D} = \mathbf{QPQ} = \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array}\right] \left[\begin{array}{cc} 1-\beta & \beta \\ \beta & 1-\beta \end{array}\right] \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1-2\beta \end{array}\right]$$

The diagonal elements of $\mathbf{D}$ are just the eigenvalues. We can now compute

$$\mathbf{P}^n = \mathbf{Q}\mathbf{D}^n\mathbf{Q} = \mathbf{Q}\left[\begin{array}{cc} 1^n & 0 \\ 0 & (1-2\beta)^n \end{array}\right]\mathbf{Q}$$

$$\begin{aligned}
\mathbf{P}^n &= \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array}\right] \left[\begin{array}{cc} 1^n & 0 \\ 0 & (1-2\beta)^n \end{array}\right] \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array}\right] \\
&= \left[\begin{array}{cc} 1/\sqrt{2} & \left(1/\sqrt{2}\right)(1-2\beta)^n \\ 1/\sqrt{2} & -\left(1/\sqrt{2}\right)(1-2\beta)^n \end{array}\right] \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{1}{2}+\frac{1}{2}(1-2\beta)^n & \frac{1}{2}-\frac{1}{2}(1-2\beta)^n \\ \frac{1}{2}-\frac{1}{2}(1-2\beta)^n & \frac{1}{2}+\frac{1}{2}(1-2\beta)^n \end{array}\right]
\end{aligned}$$

The second term in each element vanishes exponentially with increasing n. Hence,

$$\mathbf{P}^n \to \left[\begin{array}{cc} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{array}\right]$$

for large n. The probabilities at the input to stage n then approach $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n \to \left[\frac{1}{2}, \frac{1}{2}\right]$.

7. A number of communication channels are cascaded so that the output of one channel is the input of the next. Let P_k be the channel matrix of channel k. Show how to compute the channel matrix of the cascaded combination in terms of the individual channel matrices. Apply your result to the determination of the channel matrix of the cascaded BSC channels in the previous problem. Does your result agree with the conclusions of problem 6?

We have essentially done this problem for the case of the BSC in the previous problem. By the same approach, $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n$. Computing $\mathbf{P}^n$ proceeds in the same way, except that diagonalization of a non-symmetric matrix would be more difficult. If the channel matrix is symmetric then the same procedure will apply directly.