MSDM 5058 Information Science
Assignment 2 (due 23th March, 2024)

Submit your assignment solution on canvas. You may discuss with others or seek help from your TA, but should not directly copy from others. Otherwise, it will be considered as plagiarism.

(1) **Average Entropy**

Let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the entropy function of a binary source. (a) Use $\log_2 3 = 1.585$, to evaluate $H(1/3)$. (b) Calculate the average entropy $H(p)$ when the probability p is chosen uniformly in the range $0 \le p \le 1$. (*Hint: For* (*b*), *use* $\log_2 x = \frac{\ln x}{\ln 2}$, *where ln is the natural logarithm.*)

Solution:

(a)

$$\mathrm{H}\left(\frac{1}{3}\right) = -\frac{1}{3}\log_2\frac{1}{3} - \left(1-\frac{1}{3}\right)\log_2\left(1-\frac{1}{3}\right) ==> H\left(\frac{1}{3}\right) = \log_2 3 - \frac{2}{3}$$

Therefore, H(1/3) = 0.918 bits

(b)

$$H(p) = -\int_0^1 \frac{p\ln p + (1-p)\ln(1-p)}{\ln 2}dp = -\frac{2\int_0^1 x\ln x\,dx}{\ln 2}$$

$$= -\frac{2}{\ln 2}\left(\frac{x^2}{2}\ln x - \frac{x^2}{4}\right)\Bigg|_0^1 = \frac{1}{2\ln 2} = 0.721 \text{ bits}$$

(2) **Mutual Information for correlated normal distributions**

X and Y are two correlated random normal variables with the following joint probability distributions

$$\binom{X}{Y} \sim N_2\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

Evaluate $I(X;Y)$ and comment on the cases when $\rho = 1$, 0 and -1.

Solution:

$$H(X) = -\int f(x)\ln f(x)\ dx$$

For normal distribution function with $X \sim N(0, \sigma^2)$, the probability distribution is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

The entropy is

$$H(X) = -\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \left[-\frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}\right] dx = \frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2)$$

The joint probability distribution function takes the form

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} e^{-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2}}$$

$$H(X, Y) = \iint f(x.y)\ln f(x, y)\ dxdy$$

For the joint entropy, we have

$$H(X, Y) = -\iint \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} e^{-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2}} \left[-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2} - \ln\left[2\pi\sigma^2\sqrt{(1-\rho^2)}\right]\right] dx\, dy$$

Changing variables, $x \to \frac{(x-\rho y)}{\sqrt{(1-\rho^2)}}$, one can easily find the joint entropy. We now separate into two parts. For the integral involving the first term in the square bracket, we have,

$$-\iint \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} e^{-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2}} \left[-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2}\right] dx\, dy$$

$$= \iint \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} e^{-\frac{(x-\rho y)^2+y^2(1-\rho^2)}{2(1-\rho^2)\sigma^2}} \left[\frac{(x-\rho y)^2+y^2(1-\rho^2)}{2(1-\rho^2)\sigma^2}\right] dx\, dy = 1$$

For the integral involving the second term in the square bracket, we have

$$\iint \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} e^{-\frac{(x^2+y^2)-2\rho xy}{2(1-\rho^2)\sigma^2}} \ln\left[2\pi\sigma^2\sqrt{(1-\rho^2)}\right] dxdy$$

$$= \ln\left[2\pi\sigma^2\sqrt{(1-\rho^2)}\right]$$

Therefore, $H(X,Y) = 1 + \ln\left[2\pi\sigma^2\sqrt{(1-\rho^2)}\right]$

Since $I(X;Y) = H(X) + H(Y) - H(X,Y)$, by substituting the results above, $I(X;Y) = -\frac{1}{2}\ln(1-\rho^2)$

(i) $\rho = 1$: This means knowing X implies perfect knowledge about Y. Therefore, the mutual information becomes infinite, which agrees with the formula above.

(ii) $\rho = 0$: In this case, X and Y are independent of each other. Hence, the mutual information is zero, which agrees with the formula.

(iii) $\rho = -1$: This case again means knowing X implies perfect knowledge about Y. Therefore, the mutual information becomes infinite.

(3) **Channel Capacity**

Consider a binary symmetric communication channel, whose input source is the alphabet X = {0, 1} with probabilities {0.5, 0.5}; whose output alphabet is Y = {0, 1}; and whose channel matrix is

$$\begin{pmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{pmatrix}$$

Where α is the probability of transmission error when sending $X = 0$ and β is the probability of transmission error when sending $X = 1$.

(a) What is the entropy of the source, $H(X)$?

(b) What is the probability distribution of the outputs, $p(Y)$, and the entropy of this output distribution, $H(Y)$?

(c) What is the joint probability distribution for the source and the output, $p(X, Y)$, and what is the joint entropy, $H(X, Y)$?

(d) What is the mutual information of this channel, $I(X; Y)$, as a function of α and β?

(e) How many combinations of (α, β) are there for which the mutual information of this channel is maximal? What are those values, and what then is the capacity of such a channel in bits?

(f) What condition do (α, β) satisfy when the capacity of this channel is minimal? What is the channel capacity in that case?

Solution:

(a) Entropy of the source, $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$, since $p = 1/2$, therefore, $H(X)$ is 1 bit.

(b) The probability distribution of the output are:

$$p(y=0) = (0.5)(1-\alpha) + 0.5\beta$$
$$p(y=1) = 0.5\alpha + (0.5)(1-\beta)$$

The entropy of this distribution is therefore,

$$H(Y) = 1 - (0.5)(1-\alpha+\beta)\log_2(1-\alpha+\beta) - 0.5(1-\beta+\alpha)\log_2(1-\beta+\alpha)$$

(c) The joint probability distribution $p(X, Y)$ is given by

$$\begin{pmatrix} 0.5(1-\alpha) & 0.5\beta \\ 0.5\alpha & 0.5(1-\beta) \end{pmatrix}$$

The entropy of this joint distribution is

$$H(X,Y) = -\sum_{x,y} p(x,y)\log_2 p(x,y)$$
$$= 1 - 0.5\alpha\log_2(\alpha) - 0.5(1-\alpha)\log_2(1-\alpha) - 0.5\beta\log_2(\beta)$$
$$- 0.5(1-\beta)\log_2(1-\beta)$$

(d) The mutual information is given by $I(X; Y) = H(X) + H(Y) - H(X, Y)$. From the above, we have

$$I(X;Y) = 1 - 0.5(1-\alpha)\log_2\left(\frac{1+\beta-\alpha}{1-\alpha}\right) - 0.5\alpha\log_2\left(\frac{1-\beta+\alpha}{\alpha}\right)$$
$$- 0.5(1-\beta)\log_2\left(\frac{1-\beta+\alpha}{1-\beta}\right) - 0.5\beta\log_2\left(\frac{1+\beta-\alpha}{\beta}\right)$$

(e) When $(\alpha,\beta) = (0,0)$ or $(1,1)$, (corresponding to perfect transmission and perfect erroneous transmission), the mutual information reaches its maximum of 1 bit, which is also the channel capacity.

(f) When $\alpha + \beta = 1$, the channel capacity reaches its minimum, and is equal to 0.

(4) **Shannon, Fano and Huffman codes**

A source X has an alphabet of seven characters {a, b, c, d, e, f, g}, with the corresponding probability {0.01, 0.24, 0.05, 0.20, 0.47, 0.01, 0.02}.

(a) What is the entropy of X?

(b) What is the expected length of the set of codewords by using the Shannon, Fano and Huffman code, respectively?

(c) Give an example of a set of codewords for X using the Shannon code, Fano code and Huffman code respectively and explain how you get it.

(d) Which code is the optimal code for X and how much greater is its expected length than the entropy of X?

Solution:

(a) Entropy of X is: $H[X] = -\sum_i p_i \log_2 p_i = 1.932$

(b) The expected length L of a set of codewords is, $L = \sum_i p_i l_i$.

For the Shannon code, each codeword has a length of $\left\lceil \log_2 \frac{1}{p_i} \right\rceil$. This corresponds to 7, 3, 5, 3, 2, 7, 6 respectively. The expected length of the Shannon code is 2.77

The lengths of the codewords correspond to the Fano code are 6, 2, 4, 3, 1, 6, 5 respectively, and the expected length of the Fano code is 1.97

The lengths of the codewords correspond to the Huffman code are also 6, 2, 4, 3, 1, 6, 5 and the expected length of the Huffman code is 1.97

(c) For the Shannon code, an example of the set of codewords is,

{a: 1000110; b: 010; c: 10000; d: 011; e: 00; f: 1000111; g: 100010}

For the Fano code, one begins by building the tree, e.g., in the first division, we put e in the first group and the other 6 characters in the second group. One can continue with the second, third and fourth division until each character is assigned a codeword. An example of the set of codewords by using Fano's method is:

{a: 111110; b: 10; c: 1110; d: 110; e: 0; f: 111111; g: 11110}.

For the Huffman code, we start from the two characters with the smallest probability, which are a and f, and assign them with the longest codewords. Here, we assign 000000 and 000001 to a and f, respectively. We combine them to give a probability of 0.02, and continue the procedure. An example of the set of codewords is:

{a: 000000; b: 01; c: 0001; d: 001; e: 1 f: 000001; g: 00001}

(d) Both Fano and Huffman codes are optimal for X and the expected length is 1.970 – 1.932 = 0.038 greater than the entropy.