Information Theory and Coding

Outline

- Practice Problems

Formulas

- ➡ Entropy $$H(X) = \sum_{i=1}^{n} P(x_i)I(x_i)$$
- ➡ Conditional Entropy $$H(X|Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j)log(\frac{1}{P(x_i|y_j)})$$
- ➡ Joint Entropy $$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j)logP(x_i, y_j)$$
  $$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|y)$$
- ➡ Mutual Information $$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$
  $$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Example 15.1. A source produces one of four possible symbols during each interval having probabilities $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, $P(x_3) = P(x_4) = \frac{1}{8}$. Obtain the information content of each of these symbols.

Example 15.3. Calculate the amount of information if binary digits (binits) occur with equal likelihood in a binary PCM system.

Example 15.4. In a binary PCM if '0' occur with probability $\frac{1}{4}$ and '1' occur with probability equal to $\frac{3}{4}$, then calculate the amount of information carried by each binit.

Example 15.5. If there are M equally likely and independent symbols, then prove that amount of information carried by each symbol will be,

$$I(x_i) = N \text{ bits.}$$

where $M = 2^N$ and N is an integer.

Solution: Since, it is given that all the M symbols are equally likely and independent, therefore, ~~the probability of occurrence of each symbol must be~~ $\frac{1}{M}$.

We know that amount of information is given as,

$$I(x_i) = \log_2 \frac{1}{P(x_i)} \quad \ldots(i)$$

~~Here, probability of each message is,~~ $P(x_i) = \frac{1}{M}$.

Hence, equation (i) will be,

$$I(x_i) = \log_2 M \quad \ldots(ii)$$

Further, we know that $M = 2^N$, hence equation (ii) will be,

$$I(x_i) = \log_2 2^N = N \log_2 2$$

$$= N \frac{\log_{10} 2}{\log_{10} 2} = N \text{ bits}$$

Hence, amount of information carried by each symbol will be 'N' bits. We know that $M = 2^N$. This means that there are 'N' binary digits (binits) in each symbol. This indicates that when the symbols are equally likely and coded with equal number of binary digits (binits), then the information carried by each symbol (measured in bits) is numerically same as the number of binits used for each symbol.

Conditional Entropy and Mutual Information

Example 15.6. Prove the statement stated as under:

"If a receiver knows the message being transmitted, the amount of information carried will be zero."

Example 15.7. If $I(x_1)$ is the information carried by symbols x_1 and $I(x_2)$ is the information carried by message x_2, then prove that the amount of information carried compositely due to x_1 and x_2 is $I(x_1, x_2) = I(x_1) + I(x_2)$.

Example 15.9. A Discrete Memoryless Source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $P(x_1) = 0.4$, $P(x_2) = 0.3$, $P(x_3) = 0.2$, $P(x_4) = 0.1$.

(*i*) Calculate $H(X)$.

(*ii*) Find the amount of information contained in the messages $x_1\, x_2\, x_1\, x_3$ and $x_4\, x_3\, x_3\, x_2$, and compare with the $H(X)$ obtained in part (*i*). *(U.P. Tech-Semester Exam. 2002-2003)*

Example 15.10. Consider a binary memoryless source X with two symbols x_1 and x_2. Prove that $H(X)$ is maximum when both x_1 and x_2 equiprobable.

Solution: Here, let us assume that

$$P(x_1) = \alpha \text{ so that } P(x_2) = 1 - \alpha.$$

We know that entropy is given by

$$H(X) = -\sum_{i=1}^{m} P(x_i) \log_2 P(x_i) \text{ bits/symbol}$$

Thus,
$$H(X) = -\alpha \log_2 \alpha - (1-\alpha) \log_2 (1-\alpha) \quad \ldots(i)$$

Differentiating above equation with respect to α, we get

or
$$\frac{dH(X)}{d\alpha} = \frac{d}{d\alpha}[-\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)]$$

Using the relation

$$\frac{d}{dx}\log_b y = \frac{1}{y}\log_b e \frac{dy}{dx},$$

Using the relation $I(x_i) = \log_2 \frac{1}{P(x_i)}$.

We get
$$\frac{dH(X)}{d\alpha} = -\log_2 \alpha + \log_2(1-\alpha)$$
$$= \log_2 \frac{1-\alpha}{\alpha}$$

Note that the maximum value of $H(X)$ requires that
$$\frac{dH(X)}{d\alpha} = 0$$

This means that
$$\frac{1-\alpha}{\alpha} = 1 \quad \text{or} \quad \alpha = \frac{1}{2}$$

Note that $H(X) = 0$ when $\alpha = 0$ or 1.

When $P(x_1) = P(x_2) = \frac{1}{2}$, $H(X)$ is maximum and is given by
$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1 \text{ b/symbol}$$

Hence Proved

Example 15.11. Verify the following expression:

$$0 \leq H(X) \leq \log_2 m$$

where m is the size of the alphabet of X.

Solution: Proof of the lower bound:

Since $$0 \leq P(x_i) \leq 1,$$

$$\frac{1}{P(x_i)} \geq 1 \quad \text{and} \quad \log_2 \frac{1}{P(x_i)} \geq 0$$

Then, it follows that

$$P(x_i) \log_2 \frac{1}{P(x_i)} \geq 0$$

Thus, $$H(X) = \sum_{i=1}^{m} P(x_i) \log_2 \frac{1}{P(x_i)} \geq 0$$

Next, we note that

$$P(x_i) \log_2 \frac{1}{P(x_i)} = 0$$

If and only if $P(x_i) = 0$ or 1. Since

$$\sum_{i=1}^{m} P(x_i) = 1$$

when $P(x_i) = 1$ then $P(x_j) = 0$ for $j \neq i$. Thus, only in this case, $H(X) = 0$.

Proof of the upper bound

Let us consider two probability distributions $[P(x_i) = P_i]$ and $[Q(x_i) = Q_i]$ on the alphabet $\{x_i\}$, $i =$ 1, 2, ... m, such that

$$\sum_{i=1}^{m} P_i = 1 \quad \text{and} \quad \sum_{i=1}^{m} Q_i = 1 \qquad \text{...(ii)}$$

We know that

$$\sum_{i=1}^{m} P_i \log_2 \frac{Q_i}{P_i} = \frac{1}{\ln 2} \sum_{i=1}^{m} P_i \ln \frac{Q_i}{P_i}$$

Next, using the inequality

$$\ln \alpha \le \alpha - 1 \quad \alpha \ge 0 \qquad \text{...(iii)}$$

and noting that the equality holds only if $\alpha = 1$, we obtain

$$\sum_{i=1}^{m} P_i \ln \frac{Q_i}{P_i} \le \sum_{i=1}^{m} P_i \left(\frac{Q_i}{P_i} - 1 \right) = \sum_{i=1}^{m} (Q_i - P_i)$$

$$= \sum_{i=1}^{m} Q_i - \sum_{i=1}^{m} P_i = 0 \quad \text{by using equations (ii).} \qquad \text{...(iv)}$$

Hence, we have

$$\sum_{i=1}^{m} P_i \log_2 \frac{Q_i}{P_i} \le 0 \qquad \text{...(v)}$$

where the equality holds only if $Q_i = P_i$ for all i.

Setting $$Q_i = \frac{1}{m}, \qquad i = 1, 2, \ldots, m \qquad \ldots(vi)$$

We get $$\sum_{i=1}^{m} P_i \log_2 \frac{1}{P_i m} = -\sum_{i=1}^{m} P_i \log_2 P_i - \sum_{i=1}^{m} P_i \log_2 m = H(X) - \log_2 m \sum_{i=1}^{m} P_i \qquad \ldots(vii)$$

$$= H(X) - \log_2 m \leq 0$$

Therefore, we get $H(X) \leq \log_2 m$

and the quality holds only if the symbols in X are equiprobable.

Example 15.12. A high-resolution black-and-white TV picture consists of about 2×10^6 picture elements and 16 different brightness levels. Pictures are repeated at the rate of 32 per second. All picture elements are assumed to be independent, and all levels have equal likelihood of occurrence. Calculate the average rate of information conveyed by this TV picture source.

(U.P. Tech. (C.O.), 2003)

Solution: We know that entropy is given by

$$H(X) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i) \text{ bits/symbol} \quad \ldots(i)$$

Here, given that $\quad m = 16, \quad P(x_i) = \dfrac{1}{16}$

Substituting all the values in equation (i), we get

$$H(X) = -\sum_{i=1}^{16}\frac{1}{16}\log_2\frac{1}{16} = 4 \text{ bits/element}$$

The rate, r, at which symbols are generated is given by

$$r = 2(10^6)(32) = 64(10^6) \text{ elements/second}$$

Hence, the average rate of information conveyed is given by

$$R = rH(X) = 64(10^6)(4)$$

or $\quad R = 256(10^6) \text{ b/s} = 256 \text{ Mb/s} \quad$ **Ans.**

CUET

Example 15.16. A discrete source emits one of five symbols once every millisecond with probabilities 1/2, 1/4, 1/8, 1/16 and 1/16 respectively. Determine the source entropy and information rate.

Example 15.18. An analog signal bandlimited to 10 kHz is quantized in 8 levels of a PCM system with probabilities of 1/4, 1/5 1/5, 1/10, 1/10, 1/20, 1/20 and 1/20 respectively. Find the entropy and the rate of information.

Example 15.19. Given a binary channel shown in figure 15.6.

(*i*) Find the channel matrix of the channel.

(*ii*) Find $P(y_1)$ and $P(y_2)$ when $P(x_1) = P(x_2) = 0.5$.

(*iii*) Find the joint probabilities $P(x_1, y_2)$ and $P(x_2, y_1)$ when $P(x_1) = P(x_2) = 0.5$. *(PTU-2001)*

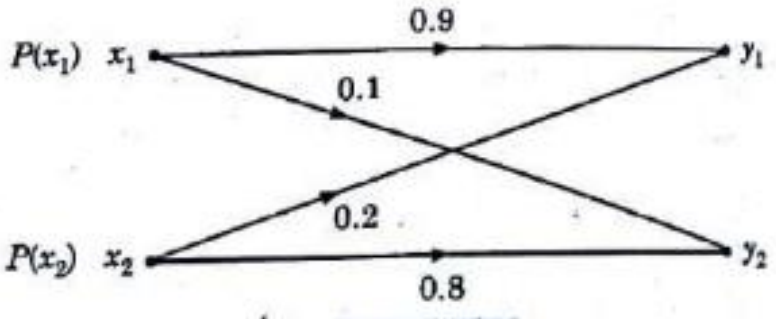


Fig. 15.6

Example 15.20. Two binary channels of problem 15.19 are connected in cascade as shown in figure 15.7

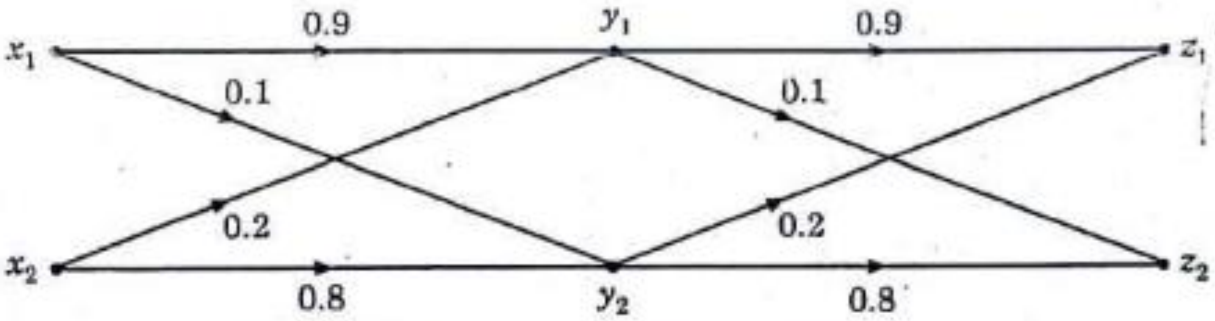


Fig. 15.7. *Two binary channels in cascade.*

(i) **Find the overall channel matrix of the resultant channel, and draw the resultant equivalent channel diagram.**

(ii) **Find $P(z_1)$ and $P(z_2)$ when $P(x_1) = P(x_2) = 0.5$**

Example 4: The international Morse code uses a sequence of symbols of dots and dashes to transmit letters of the English alphabet. The dash is represented by a current pulse of duration 2 ms and dot by duration of 1 ms. The probability of dash is half as that of dot. Consider 1ms duration of gap is given in between the symbols. Calculate

a) Self information of a dot and a dash
b) Average information content of a dot-dash code
c) Average rate of information

Solution

a) Let p_{dot} and p_{dash} be the probabilities of dot and dash, respectively. Given

$$p_{dash} = \frac{1}{2}p_{dot}$$

Also $p_{dot} + p_{dash} = 1$. Therefore,

$$p_{dot} + \frac{1}{2}p_{dot} = 1 \Rightarrow p_{dot} = \frac{2}{3}$$

We can also deduce that

$$p_{dash} = \frac{1}{2} p_{dot} = \frac{1}{3}$$

Now,

$$I(dot) = -\log_2 p_{dot} = -\log_2 \frac{2}{3} = 0.5849 \; bits$$

$$I(dash) = -\log_2 p_{dash} = -\log_2 \frac{1}{3} = 1.5849 \; bits$$

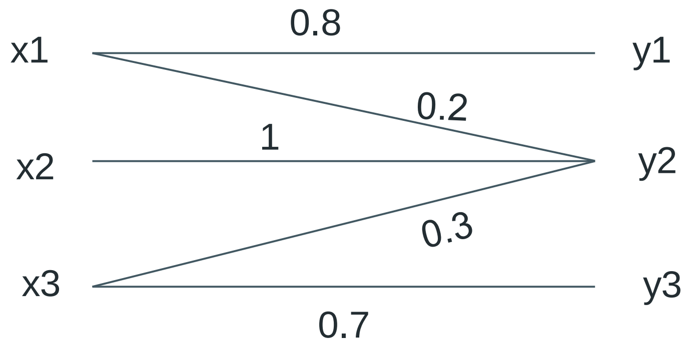
b) $H(S) = -(p_{dot} \log_2 p_{dot} + p_{dash} \log_2 p_{dash}) = 0.9182 \; bits/sym$

c) From the probabilities of dot and dash, it is clear that for every three symbols transmitted, there will be one symbol of type dash and two of type dot. Also the duration of dash is 2ms and that of dot is 1ms, and the 1ms gap is left in between the symbols. Therefore a total of (1+1+1+1+2+1) = 7ms time is required to transmit an average of 3 symbols ($\cdot \cdot -$). So the symbol rate is given by

$$r_s = \frac{\frac{3}{7} sym}{ms} = \frac{\frac{3000}{7} sym}{s}$$

$$\Rightarrow R_s = H(S) \times r_s = 0.9182 \times \frac{3000}{7} = 393.51 \; bits/s$$

Exercise 1: Given that $P[X] = [0.3\ 0.4\ 0.3]$. Find all entropies.



Exercise 2: Given JPM.

$$P(X,Y) = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.1 & 0.3 & 0 & 0 \\ 0 & 0.05 & 0.10 & 0 \\ 0 & 0 & 0.05 & 0.1 \\ 0 & 0 & 0.05 & 0 \end{bmatrix}$$

Find all entropies.

Exercise 1: A black and white TV picture consists of **525** lines of picture information. Assume that each line consists of **525** picture elements (pixels) and that each element can have **256** brightness level. Pictures are repeated at a rate of **30 frames/sec**. Calculate the average rate of information conveyed by a TV set to a viewer.

Exercise 2 : In a facsimile transmission of there are about **2.25 x 10^6** pixel per frame. For a good reproduction **12 brightness levels** are necessary. Assume all these levels are equally likely to occur. Find the rate of information transmission if one picture is to be transmitted in every 3 minute.