INTRODUCTION TO CODING THEORY

Multiple Choice Type Questions

1. Entropy is basically a measure of [WBUT 2013, 2014, 2017]
a) rate of information
b) average information
c) probability of information
d) disorder of information

Answer: (d)

2. The entropy of a message source generation four messages with probabilities 0.5, 0.25, 0.125 is [WBUT 2015]
a) 1.0 bit/message
b) 1.75 bit/message
c) 3.32 bit/message
d) 5.93 bit/message

Answer: 1.375 bit/message

3. Higher degree of uncertainty means [WBUT 2015]
a) lesser information
b) more information
c) zero information
d) none of these

Answer: (b)

4. Which of the following is not of information? [MODEL QUESTION]
a) bit
b) decit
c) Hz
d) nat

Answer: (c)

5. Channel capacity is exactly equal to [MODEL QUESTION]
a) bandwidth of demand
b) amount of information per second
c) noise rate in the demand
d) none of these

Answer: (b)

6. If a source produces five symbols with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{16}$, then the source entropy $H(x)$ is [MODEL QUESTION]
a) $3b/$ symbols
b) $5.5b/$ symbols
c) $2.875b/$ symbols
d) $1.875b/$ symbols

Answer: (d).

7. Mutual information of a channel with independent input and output is [MODEL QUESTION]
a) Zero
b) Constant
c) Infinite
d) Variable

Answer: (a)

8. Information content in a universally true event is [MODEL QUESTION]
a) Infinite
b) Zero
c) Positive content
d) Negative content

Answer: (b)

9. The main purpose of channel coding is [MODEL QUESTION]
a) Maximizing the efficiency of communication
b) Maximizing the reliability of communication
c) Decreasing redundancy during coding
d) Maximizing the S/N ratio

Answer: (b)

10. The channel capacity is a measure of [MODEL QUESTION]
a) Entropy rate
b) Maximum rate of information a channel can handle
c) Information contents of messages transmitted in a channel

Answer: (b)

Short Answer Type Questions

1. A source is emitting 4 symbols with probabilities 1/2, 1/4, 1/8 and 1/8. What is entropy of sources and what should be code length if code efficiency is 100%. **[WBUT 2013]**

Answer:

$$S_1 = \frac{1}{2}, S_2 = \frac{1}{4}, S_3 = \frac{1}{8}, S_4 = \frac{1}{8}$$

$$H(x) = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + 2\times\frac{1}{8}\log_2(8)$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} = 1.375 \text{ bits.}$$

2. What is Shannon-Hartley channel capacity theorem? **[WBUT 2015]**

Answer:

Hartley Shannon Law states that for a Gaussian Noisy Channel, the channel capacity is given by $C = B\log_2\left[1+\frac{S}{N}\right]$ bits/sec.

where B = Channel bandwidth
S = Average signal power
N = Average noise power

If $\frac{N_0}{2}$ is the power spectral density of noise in watts / Hz, then N = N_0 B and

$$C = B\log_2\left[1+\frac{S}{N_0 B}\right] \text{bits/sec.}$$

3. a) What do you mean by Entropy and Information rate? **[WBUT 2018]**
b) What is Shannon-Fano algorithm?

Answer:

a) 1st Part: *Refer to Question No. 1(a) of Long Answer Type Questions.*
2nd part: *Refer to Question No. 5 of Short Answer Type Questions.*

b) *Refer to Question No. 4(b) of Long Answer Type Questions.*

4. a) Define and explain the term "Channel Capacity".
b) Calculate the capacity of a channel with bandwidth 2900 Hz and signal to noise ratio of 316.2. **[MODEL QUESTION]**

Answer:

a) The number of bits of information that a channel can transmit per unit if time is called channel capacity. Mathematically, channel capacity,

$$C = \underset{T \to \infty}{Lt} \frac{1}{T} \log_2 N(T)$$

Where, N(T) = Number of allowed signal sequences in a duration T.

b) Given that,

B = 2900 H_z. $\qquad S/N = 316.2$

$$\therefore C = B\log_2\left(1+\frac{S}{N}\right) = 2900 \log_2(1+316.2) = 7.5 \text{ bits/symbol}.$$

5. What do you mean by rate of information? **[MODEL QUESTION]**

Answer:

If a message source having entropy '*H*' generates messages at the rate of '*r*' messages per second, then the rate of information '*R*' is defined as the average number of bits of information per second.

Then

$$R = \frac{\text{Average number of information}}{\text{Second}}$$

$$= \frac{\text{Average number information}}{\text{Number of messages}} \times \frac{\text{Number of messages}}{\text{Second}}$$

$$= H \times r$$

where H = Entropy

Thus $R = rH$ bits per second

6. What are the five entropies associated with a digital communication channel? What are their significances? **[MODEL QUESTION]**

Answer:

The five entropies associated with a digital communication channel which is a two-dimensional probability scheme are the following:

H(X), H(Y), H(XY), H(X/Y) and H(Y/X)

Let X represents a transmitter and Y represents a receiver and information goes from the transmitter to the receiver through a noisy channel.

Here H(X) = Average information per character at the transmitter = Entropy of the transmitter

H(Y) = Average information per character at the receiver = entropy of the receiver.

H(XY) = Average information per pair of the transmitted and received characters.

= Average uncertainty of the communication system as a whole.

H(X/Y) = Conditional Entropy giving a measure of information about the transmitter where it is known that X is transmitted

H(X/Y) indicates how well one can recover the transmitter symbols from the received symbols. Thus it gives a measure of equivocation.

H(Y/X) indicates how well one can recover the received symbols from the transmitted symbols. It gives a measure of error or noise.

Long Answer Type Questions

1. a) What do you mean by entropy?

b) Consider a Binary memory-less source X with two symbols $x1$ and $x2$. Prove that $H(X)$ is maximum when both $x1$ and $x2$ are equiprobable.

c) The parity check matrix of a particular (7,4) linear block code is given by:

$$[H]=\begin{vmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{vmatrix}$$

i) Find the Generator Matrix

ii) List all the Code vectors

iii) What is the minimum distance between the code vectors?

iv) How many errors can be detected?

v) How many errors can be corrected? **[WBUT 2013]**

Answer:

a) Entropy:

Definition

The average information per message of a source is called source entropy or simply entropy. It is denoted by H and

$$H=-\sum_{1}^{m} p_i \log p_i \text{ binits} = \sum_{i=1}^{m} p_i \log\frac{1}{p_i} \text{ binits} = \sum_{i=1}^{m} p_i I_i \text{ binits}$$

where m is the total number of messages in the source and p_i is the probability of occurrence of the i^{th} message. I_i is the information of the i^{th} message. Note that $\sum_{i=1}^{m} p_i = 1$.

Properties of entropy

1. If all the probabilities of messages except one in a source are zero, the entropy $H(x) = 0$. This is the lower bound of the entropy.
2. If all the messages in a source are equiprobable, then the entropy $H(x) = \log_2 K$ where K is the radix or number of symbols of the alphabet of the source. This is the upper bound of the entropy.
3. The entropy of a source is bounded as $0 \le H(x) < \log_2 K$.
4. For a binary system, maximum entropy occurs when $p = \frac{1}{2}$.

b) Let $P(x_1) = \alpha$ $P(x_2) = 1-\alpha$

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

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

Using the relation

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

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

The maximum value of H(X) requires that

$$\frac{dH(X)}{d\alpha} = 0$$

that is,

$$\frac{1-\alpha}{\alpha} = 1 \rightarrow \alpha = \frac{1}{2}$$

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

$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1 \text{ b/symbol} \quad \ldots (2)$$

c) Here $n = 7$ and $k = 4$

$\therefore$ Number of check bits are $n-k = 7-4$ i.e. $q = 3$

Thus $n = 2^q - 1 = 2^3 - 1 = 7$

This shows that the given code is hamming code.

To determine the P sub-matrix:

The parity check matrix is of $q \times n$ size and is given by following. It can be written as, (with $q = 3$ and $n = 7$ and $k = 4$)

$$[H]_{3\times 7} = \begin{bmatrix} P_{11} & P_{21} & P_{31} & P_{41} & \cdots & 1 & 0 & 0 \\ P_{12} & P_{22} & P_{32} & P_{42} & \cdots & 0 & 1 & 0 \\ P_{13} & P_{23} & P_{33} & P_{43} & \cdots & 0 & 0 & 1 \end{bmatrix} \quad \ldots (1)$$

$$= \left[P^T : I_3\right]$$

On comparing parity check matrices of equation we get,

$$P^T = \begin{bmatrix} P_{11} & P_{21} & P_{31} & P_{41} \\ P_{12} & P_{22} & P_{32} & P_{42} \\ P_{13} & P_{23} & P_{33} & P_{43} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

Therefore the P submatrix can be obtained as,

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \\ P_{41} & P_{42} & P_{43} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}_{4\times 3} \quad \ldots (2)$$

i) To obtain the generator matrix (G):

The generator matrix G is given as,

$$G = \left[I_k : P_{k\times q}\right]_{k\times n}$$

with k = 4, q = 3 and n = 7 the above equation becomes,

$$G = \left[I_4 : P_{4\times 3}\right]_{4\times 7}$$

Putting the identity matrix I_4 of size 4 × 4 and parity sub-matrix $P_{4\times 3}$ of size 4 × 3 as obtained

$$G = \begin{bmatrix} \underbrace{\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}}_{I_{4\times 4}} & \begin{matrix} : \\ : \\ : \\ : \end{matrix} & \underbrace{\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix}}_{P_{4\times 3}} \end{bmatrix}_{4\times 7} \quad \ldots (3)$$

This is the required generator matrix

ii) To find all the code words:

To obtain equations for check bits

The check bits can be obtained i.e.,

$$C = MP$$

In the more general form with $q = 3, k = 4$)

$$[C_1 C_2 C_3]_{1\times 3}] = [m_1 \quad m_2 \quad m_3 \quad m_4]_{1\times 4} [P]_{4\times 3}$$

$$[C_1C_2C_3]=[m_1 \quad m_2 \quad m_3 \quad m_4]\begin{bmatrix}1 & 1 & 1\\1 & 1 & 0\\1 & 0 & 1\\0 & 1 & 1\end{bmatrix}_{4\times 3}$$

Solving the above equation with mod-2 addition we get,

$$C_1=(1\times m_1)\oplus(1\times m_2)\oplus(1\times m_3)\oplus(0\times m_4)$$

$$C_2=(1\times m_1)\oplus(1\times m_2)\oplus(0\times m_3)\oplus(1\times m_4)$$

and $$C_3=(1\times m_1)\oplus(0\times m_2)\oplus(1\times m_3)\oplus(1\times m_4)$$

Thus the above equations are,

$$\left.\begin{aligned}C_1&=m_1\oplus m_2\oplus m_3\\C_2&=m_1\oplus m_2\oplus m_4\\ \text{and}\quad C_3&=m_1\oplus m_3\oplus m_4\end{aligned}\right\}\quad ...(4)$$

To determine the code vectors

Consider for example $(m_1 \quad m_2 \quad m_3 \quad m_4)=1\,0\,1\,1$ we get,

$$C_1=1\oplus 0\oplus 1=0$$

$$C_2=1\oplus 0\oplus 1=0$$

and $$C_3=1\oplus 1\oplus 1=1$$

Thus for message vector of (1 0 1 1) the check bits are $(C_1C_2C_3)=001$. Therefore, the systematic block code of the code vector (codeword) can be written as,

$$(m_1m_2m_3m_4C_1C_2C_3)=(1\ 0\ 1\ 1:0\ 0\ 1)$$

Using the same procedure as given above we can obtain the other codeword or code vectors. Table 1 lists all the code vectors (codeword). Table also lists the weight of each codeword.

Weight of each codeword.

Sl. No.	Message vector M				Check bits (C)			Code vector or codeword X							Weight of Code Vectorw(x)
	m_1	m_2	m_3	m_4	C_1	C_2	C_3	m_1	m_2	m_3	m_4	C_1	C_2	C_3	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	1	1	0	0	0	1	0	1	1	3
3	0	0	1	0	1	0	1	0	0	1	0	1	0	1	3
4	0	0	1	1	1	1	0	0	0	1	1	1	1	0	4
5	0	1	0	0	1	1	0	0	1	0	0	1	1	0	3
6	0	1	0	1	1	0	1	0	1	0	1	1	0	1	4

Sl. No.	Message vector M				Check bits (C)			Code vector or codeword X							Weight of Code Vectorw(x)
	m_1	m_2	m_3	m_4	C_1	C_2	C_3	m_1	m_2	m_3	m_4	C_1	C_2	C_3	
7	0	1	1	0	0	1	1	0	1	1	0	0	1	1	4
8	0	1	1	1	0	0	0	0	1	1	1	0	0	0	3
9	1	0	0	0	1	1	1	1	0	0	0	1	1	1	4
10	1	0	0	1	1	0	0	1	0	0	1	1	0	0	3
11	1	0	1	0	0	1	0	1	0	1	0	0	1	0	3
12	1	0	1	1	0	0	1	1	0	1	1	0	0	1	4
13	1	1	0	0	0	0	1	1	1	0	0	0	0	1	3
14	1	1	0	1	0	1	0	1	1	0	1	0	1	0	4
15	1	1	1	0	1	0	0	1	1	1	0	1	0	0	4
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7

iii) Minimum distance between code vectors

The table lists $2^k = 2^4 = 16$ code vectors along with their weights. the smallest weight of any non-zero code vector is 3. We know that the minimum distance is $d_{min} = 3$. Therefore we can write:

The minimum distance of a linear block code is equal to the minimum weight of any non zero code vector i.e.

$$d_{min} = [w(X)]_{min}; X \neq (00....0) \qquad \text{... (5)}$$

iv) & v) Error detection and correction capabilities

Since $d_{min} = 3$,

$$d_{min} \geq s + 1$$

$$3 \geq s + 1$$

or $\quad s \leq 2$

Thus two errors will be detected.

and $\quad d_{min} \geq 2t + 1$

$$3 \geq 2t + 1$$

or $\quad t \leq 1$

Thus one error will be corrected.

The hamming code ($d_{min} = 3$) always two errors can be detected and single error can be corrected by its property.

2. A DMS X has five symbols x_1, x_2, x_3, x_4 and x_5 with $P(x_1) = 0.4$, $P(x_2) = 0.19$, $P(x_3) = 0.16$, $P(x_4) = 0.15$ and $P(x_5) = 0.1$. Construct a Shannon-Fano code for X and calculate the efficiency of the code. **[WBUT 2014, 2018]**

Answer:

Message	Probability	Step 1	Step 2	Step 3	Code	Code Length
x_1	0.4	0	0		00	2
x_2	0.19	0	1		01	2
x_3	0.16	1	0		10	2
x_4	0.15	1	1	0	110	3
x_5	0.1	1	1	1	111	3

Thus the codes formed are $c_1 = 00$, $c_2 = 01$, $c_3 = 10$, $c_4 = 110$ and $c_5 = 111$

Average code length,

$$\bar{L} = (0.4\times2) + (0.19\times2) + (0.16\times2) + (0.15\times3) + (0.1\times3)$$
$$= 2.25 \text{ symbols/message}$$

Source Entropy, $H(X) = -[0.4 \log 0.4 + 0.19 \log 0.19 + 0.16 \log 0.16 + 0.15 \log 0.15 + 0.1 \log 0.1]$

$$= 2.15 \text{ bits/message.}$$

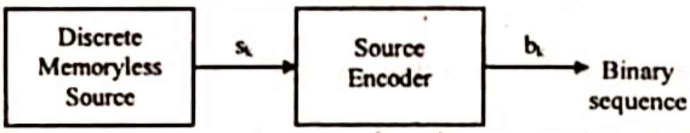
Hence efficiency $= \eta = \dfrac{H(X)}{\bar{L}} = \dfrac{2.15}{2.25} = 0.956 = 95.6\%$

3. State source coding theorem. **[WBUT 2015]**

Answer:

Shannon's source coding theorem:

Let us consider a source encoder as shown below:



The output of the discrete memoryless source is s_k which is converted to binary sequence b_k. Let the source alphabet be given by

$$s = \{s_0, s_1, \cdots\cdots, s_k\}$$

Let the corresponding probabilities be

$\{p_0, p_1, \cdots\cdots, p_k\}$ and

Code length be

$\{\ell_0, \ell_1, \cdots\cdots, \ell_k\}$ respectively

Thus the average code length i.e., the average number of bits per symbol of the source is defined as

$$\bar{L} = \sum_{k=0}^{k-1} p_k \ell_k$$

Now the Shannon's Source Coding Theorem is stated as follows:

Given a discrete memoryless source of entropy $H(s)$, the average code-word length $\bar{L}$ for any distortionless source coding is bounded as

$$\bar{L} \ge H(s)$$

The source coding theorem is also known as the "noiseless coding theorem" and "Shannon's first theorem".

If $\bar{L}_{min}$ denotes the minimum possible value of $\bar{L}$, then we define the coding efficiency of the source encoder as

$$\eta = \frac{\bar{L}_{min}}{\bar{L}}$$

For an efficient code, η approaches unity

According to source coding theorem, the minimum value of $\bar{L}$ is $H(s)$.

Hence, $\eta = \frac{H(s)}{\bar{L}}$.

Shannon's source coding theorem provides the mathematical tool for assessing data compaction of data generated by a discrete memoryless source.

4. a) What is hamming code? Write down properties of hamming distance.

[WBUT 2015]

Answer:

1st Part:

A Hamming code is an (n, k) linear block code with $r \geq 3$ check bits and $n = 2^r - 1$ and $k = n - r$

The code rate of a Hamming code is

$$R_c = \frac{k}{n} = \frac{n-r}{2^r - 1} = \frac{n}{2r-1} - \frac{r}{2^r - 1} = \frac{2^r - 1}{2^r - 1} - \frac{r}{2^r - 1} = 1 - \frac{r}{2^r - 1}$$

If $r >> 1,\ R_c \approx 1$

The minimum distance is fixed at $d_{min} = 3$. Thus a Hamming code can be used for single error correction or double error detection. Hamming codes are perfect codes and they can be binary as well as non-binary. Hamming single-error correcting codes are BCH (or Bose-Chaudhuri-Hocquenghem) codes. For single-error correcting Hamming code the minimum size n for the code word can be determined from the relation $n \geq k + \log_2(1+n)$.

Properties

It has the following properties:

1. Reflexivity: $H_d(x,y) = 0$ if and only if $x = y$.
2. Symmetry: $H_d(x,y) = H_d(y,x)$
3. Triangle inequality: $H_d(x,y) + H_d(y,z) \geq H_d(x,z)$

The metric properties of the Hamming distance allow us to use the geometry of the codespace to reason about the codes. As an example, consider the codespace $\{0, 1\}^3$ illustrated by a cube shown in Fig. 5.1. The codewords {000, 011, 101, 110} are marked

with large solid dots. It is easy to see that the Hamming distance satisfies the metric properties listed above,
e.g., $H_d(000,011)+H_d(011,111)=2+1=3=H_d(000,111)$

b) Write down Shannon-Fano algorithm. **[WBUT 2015]**

Answer:

Shannon-Fano Coding:

It is an efficient source coding technique. The algorithm for constructing Shannon – Fano codes is as follows.

Algorithm

Step 1: The messages are first arranged in the order of decreasing probabilities.

Step 2: The message set is partitioned into two most equiprobable subsets $\{x_1\}$ and $\{x_2\}$.

Step 3: A '0' is assigned to each message in one subset say $\{x_1\}$ and a '1' is assigned to each message in the other subset say $\{x_2\}$.

Step 4: The above procedures are repeated for the subset $\{x_1\}$ and $\{x_2\}$. Thus $\{x_1\}$ will be partitioned into two subsets say $\{x_{11}\}$ and $\{x_{12}\}$ and $\{x_2\}$ set will be partitioned into two subsets say $\{x_{21}\}$ and $\{x_{22}\}$.

Step 5: The code words in subset $\{x_{11}\}$ will start with 00 and in $\{x_{12}\}$ with 01. Subset $\{x_{21}\}$ will start with 10 and $\{x_{22}\}$ will start with 11.

Step 6: The procedure is continued until each subset contains only one message.

Example of Shannon-Fano Coding

Let $[X] = \{x_1, x_2, x_3, \ldots\ldots, x_8\}$ and probability

$$[P]=\left\{\frac{1}{4},\ \frac{1}{8},\ \frac{1}{16},\ \frac{1}{16},\ \frac{1}{16},\ \frac{1}{4},\ \frac{1}{16},\ \frac{1}{8}\right\}$$

Let us form the binary code words using Shannon Fano Coding.

Solution:

First Step

Message	*Probability*
x_1	1/4 = 0.25
x_6	1/4 = 0.25
x_2	1/8 = 0.125
x_8	1/8 = 0.125
x_3	1/16 = 0.0625
x_4	1/16 = 0.0625
x_5	1/16 = 0.0625
x_7	1/16 = 0.0625

Second Step

$[X_1] = [x_1, x_6]$
$[X_2] = [x_2, x_8, x_3, x_4, x_5, x_7]$

Message	Probability	Encoded Message	Subset
x_1	0.25	0	subset$\{x_1\}$
x_6	0.25	0	
x_2	0.125	1	
x_8	0.125	1	subset$\{x_2\}$
x_3	0.0625	1	
x_4	0.0625	1	
x_5	0.0625	1	
x_7	0.0625	1	

Third Step

x_1	0.25	0 0	$\{x_{11}\}$
x_6	0.25	0 1	$\{x_{12}\}$
x_2	0.125	1 0	$\{x_{21}\}$
x_8	0.125	1 0	$\{x_{21}\}$
x_3	0.0625	1 1	$\{x_{22}\}$
x_4	0.0625	1 1	$\{x_{22}\}$
x_5	0.0625	1 1	$\{x_{22}\}$
x_7	0.0625	1 1	$\{x_{22}\}$

Fourth Step

x_2	1 0 0
x_8	1 0 1
x_3	1 1 0 0
x_4	1 1 0 1
x_5	1 1 1 0
x_7	1 1 1 1

Thus the codes formed by Shannon Fano Coding techniques are c_1 = 00, c_2 = 100, c_3 = 1100, c_4 = 110 1, c_5 =1110, c_6 = 01, c_7 = 1111 and c_8 = 101.

5. a) A discrete source emits one of the five symbols once every millisecond with probabilities 1/2, 1/4, 1/8, 1/16, $\frac{1}{16}$ respectively. Determine the source entropy and information rate.

b) A DMS *X* has five symbols x_1, x_2, x_3, x_4 and x_5 with $P(x_1)=0.30$, $P(x_2)=0.20$, $P(x_3)=0.25$, $P(x_4)=0.12, P(x_5)=0.05$ and $P(x_6)=0.08$. Construct a Huffman code for *X* and calculate the efficiency of the code.

c) What do you mean by mutual information? **[WBUT 2016]**

Answer:

a) $H(s)=\sum_{i=1}^{5} p_i \log_2 \frac{1}{p_i}$

$$=\frac{1}{2}\log_2(2)+\frac{1}{4}\log_2(4)+\frac{1}{8}\log_2(8)+\frac{1}{16}\log_2(16)+\frac{1}{16}\log_2(16)$$

$$= 0.5 + 0.5 + 0.375 + 0.25 + 0.25 = 1.875 \text{ bits/symbol}$$

Information rate, R

$$R = r_s H(s) \text{ bits/sec} = 1000 \times 1.875 \text{ bits/sec.}$$

b) $H(X) = -\sum_{i=1}^{6} P(x_i) \log_2 \cdot P(x_i)$

$$= -0.3\log_2 0.3 - 0.2\log_2 0.2 - 0.25\log_2 0.25$$
$$-0.12\log_2 0.12 - 0.05\log_2 0.05 - 0.08\log_2 (0.08)$$
$$= -0.3(-1.2) - 0.2(-1.6) - 0.25(-1.38) - 0.12(2.12)$$
$$-0.05(2.99) - 0.08(2.52)$$
$$= 0.36 + 0.32 + 0.345 + 0.2544 + 0.14 + 0.20 = 1.6914$$

$$L = \sum_{i=1}^{6} P(x_i) n_i = 0.3(2) + 0.2(2) + 0.25(2) + 0.12(3) + 0.05(3) + 0.08(3)$$
$$= 0.6 + 0.4 + 0.5 + 0.36 + 0.15 + 0.24 = 2.25$$
$$\eta = \frac{H(X)}{L} = \frac{1.6914}{2.25} = 0.7517 = 75.17\%$$

c) Mutual Information:

Let us study the transfer of information from a transmitter through a channel to a receiver.

Prior to the reception of a message, the state of knowledge at the receiver about a transmitted signal x_j is the a-priori probability $p(x_j)$.

After the reception and selection of the symbol y_k, the state of knowledge concerning x_j is the conditional probability $p(x_j/y_k)$. This is called a posteriori probability.

Thus before y_K is received, the uncertainty and hence information is $-\log\lfloor p(x_j)\rfloor$ and after y_k is received the uncertainty and hence information becomes $-\log\lfloor p(x_j / y_K)\rfloor$.

Obviously, the information gained about x_j by the reception of y_K is the net reduction of its uncertainty and is known as mutual information denoted by $I(x_j; y_k)$.

Thus $I(x_j; y_K) = -\log p(x_j) + \log p(x_j/y_k) = \log \dfrac{p(x_j / y_K)}{p(x_j)}$

Mutual information is also called transferred information or trans-information.

Mutual information is symmetrical in x_j and y_k. That is

$$I(x_j; y_k) = I(y_k; x_j)$$

Self information may be treated as a special case of mutual information where $y_k = x_j$.

Then $I(x_j; x_j) = \log \dfrac{p(x_j / x_j)}{p(x_j)} = \log \dfrac{1}{p(x_j)} = I(x_j)$.

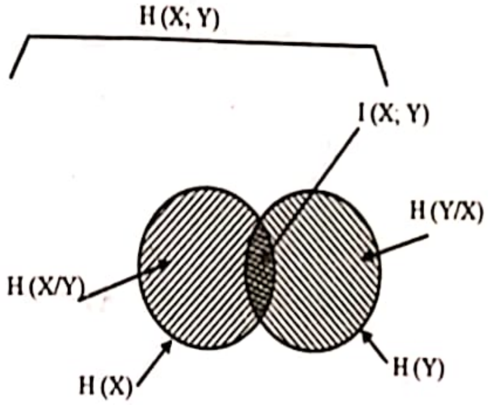
The average of mutual information i.e. the entropy corresponding to mutual information is denoted by I(X;Y). It can be shown that

$$I(X ; Y) \geq 0$$

$$I(X ; Y) = H(X) - H(X/Y)$$
$$= H(X) + H(Y) - H(X, Y)$$
$$= H(Y) - H(Y/X)$$

$$I(X; Y) = I(Y ; X)$$

The various entropies of a communication network can be represented diagrammatically by the following illustration.



Here H(X) = entropy of the channel input X. This is represented by the entire circle on the left.
H(Y) = Entropy of the channel output Y.
This is represented by the entire circle on the right.
The mutual information I(X ; Y) is represented by the overlap between the two circles.
H(X ; Y) is the joint entropy of the channel.

6. a) Explain the term entropy and information. Prove that **[WBUT 2017]**

$$H(x) = \sum P(x_1) \log\left[\frac{1}{P(x_1)}\right].$$

Answer:

Entropy: ***Refer to Question No. 1(a) of Long Answer Type Questions.***
Information: ***Refer to Question No. 5 of Short Answer Type Questions.***

Prove:

We know from kraft in equality any set of codewords of lengths ℓ_i is decodable if it satisfies, $\sum_i 2^{-\ell_i} \leq 1$

We also know that if $\sum_i 2^{-\ell_i} > 1$, no uniquely decodable code exists with those codeword lengths.

If a source generates symbols X independently with probabilities p_i, the expected codeword length per symbol is

$$L = \sum_i p_i \ell_i$$

Let $K = \sum_i 2^{-\ell_i}$.

Assume $K \le 1$.

Now, we can find a lower bound on L in terms of the p_i:

$$\sum_i p_i \ell_i \ge \sum_i p_i (\ell_i + \log K)$$

$$= \sum_i p_i \log(2^{\ell_i} K) = \sum_i p_i \log\left(\frac{2^{\ell_i} K_{p_i}}{p_i}\right) = \sum_i p_i \log\frac{1}{p_i} + \sum_i p_i \log(2^{\ell_i} K_{p_i})$$

$$\therefore \quad \sum_i p_i \ell_i + \sum_i p_i \log\left(\frac{1}{2^{\ell_i} K_{p_i}}\right) \ge \sum_i p_i \log\frac{1}{p_i}$$

Using the fact that $\log(x) \le \alpha(x-1)$

We can show that the second term on the left is always negative:

$$\sum_i p_i \log\left(\frac{1}{2^{\ell_i} K_{p_i}}\right) \le \alpha \sum_i p_i \left(\frac{1}{2^{\ell_i} K_{p_i}} - 1\right) = \alpha \sum_i \frac{1}{2^{\ell_i} K} - \alpha \sum_i p_i = \alpha(i-1) = 0$$

Thus we can show that, $L = \sum_i p_i \ell_i \ge \sum_i p_i \log\frac{1}{p_i} = H$

The quantity $H(X) = \sum_i p_i \log\frac{1}{p_i}$ is called entropy of the distribution $P(X = a_i) = p_i$ and is a fundamental quantity in the study of information theory.

b) Find the entropy of a source that produces 4 symbols A, B, C and D with probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}$. Also find the rate of information if there are 16 outcomes per second. **[WBUT 2017]**

Answer:

We know, entropy $H(X)$ is given by,

$$H(X) = \sum_i P(X_i) \log_2 \frac{1}{P(X_i)}$$

$$= \frac{1}{6}\log_2 6 + \frac{1}{3}\log_2 3 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4$$

$$= \frac{1}{6} \times 2.585 + \frac{1}{3} \times 1.585 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2$$

$$= 0.431 + 0.528 + 1 = 1.959 \text{ bits/symbol}$$

Information Rate, $R = rH$

where, r = rate of generation of message.

Here, $r = 16$ outcomes/second

So, $R = 16 \times 1.959 = 31.34$ bits/second.

c) Develop Shanon Fano code for five messages given by probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}$. Calculate the average number of bits/message. [WBUT 2017]

Answer:

$P(X_1) = \frac{1}{2}$, $P(X_2) = \frac{1}{4}$, $P(X_3) = \frac{1}{8}$, $P(X_4) = \frac{1}{16}$, $P(X_5) = \frac{1}{16}$

Symbols	Probability	Encoded message			
x_1	$\frac{1}{2}$	0			
x_2	$\frac{1}{4}$	1	0		
x_3	$\frac{1}{8}$	1	1	0	
x_4	$\frac{1}{16}$	1	1	1	0
x_5	$\frac{1}{16}$	1	1	1	1

Hence, $C_1 = 0,\ C_2 = 10,\ C_3 = 110,\ C_4 = 1110,\ C_5 = 1111$

Average number of bits/message

$$\overline{L} = 1\times\frac{1}{2} + 2\times\frac{1}{4} + 3\times\frac{1}{8} + 4\times\frac{1}{16} + 4\times\frac{1}{16}$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{1}{4} = 1 + \frac{1}{2} + \frac{3}{8} = \frac{3}{2} + \frac{3}{8} = \frac{15}{8} \text{ bits/message}$$

7. Write short notes on any *three* of the following:
a) Huffman Code [WBUT 2013]
b) Entropy [WBUT 2015]
c) Channel capacity [WBUT 2015]

Answer:

a) Huffman coding was developed by DA Huffman in the year 1952. Huffman codes are optimal codes that map one symbol to one code word. In Huffman coding, it is assumed that each pixel intensity has associated with it a certain probability of occurrence, and this probability is spatially invariant. Huffman coding assigns a binary code to each intensity value, with shorter codes going to intensities with higher probability. If the probabilities can be estimated apriori then the table of Huffman codes can be fixed in both the encoder and the decoder. Otherwise, the coding table must be sent to the decoder along with the compressed image data.

The parameters involved in Huffman coding are as follows:

- Entropy

- Average length
- Efficiency
- Variance

Otherwise, the coding table must be sent to the decoder along with the compressed image data.

Prefix Code

A code is a prefix code if no code word is the prefix of another code word. The main advantage of a prefix code is that it is uniquely decodable. An example of a prefix code is the Huffman code.

Types of Huffman Codes

Huffman codes can be broadly classified in to (i) binary Huffman code, (ii) non-binary Huffman code, (iii) extended Huffman code, and (iv) adaptive Huffman code.

b) The average information per message of a source is called source entropy or simply entropy. It is denoted by H and

$$H = -\sum_{1}^{m} p_i \log p_i \text{ binits} = \sum_{i=1}^{m} p_i \log \frac{1}{p_i} \text{ binits} = \sum_{i=1}^{m} p_i I_i \text{ binits}$$

where m is the total number of messages in the source and p_i is the probability of occurrence of the i^{th} message. I_i is the information of the i^{th} message. Note that $\sum_{i=1}^{m} p_i = 1$.

Properties of entropy

1. If all the probabilities of messages except one in a source are zero, the entropy $H(x) = 0$. This is the lower bound of the entropy.
2. If all the messages in a source are equiprobable, then the entropy $H(x) = \log_2 K$ where K is the radix or number of symbols of the alphabet of the source. This is the upper bound of the entropy.
3. The entropy of a source is bounded as $0 \le H(x) < \log_2 K$.
4. For a binary system, maximum entropy occurs when $p = \frac{1}{2}$.

c) The number of bits of information that a channel can transmit per unit if time is called channel capacity.

Channel capacity is the tightest upper bound on the amount of information that can be reliably transmitted over a communications channel. The channel capacity of a given channel is the limiting information rate (in units of information per unit time) that can be achieved with arbitrarily small error probability.

Mathematically, channel capacity,

$$C = \underset{T\to\infty}{Lt} \frac{1}{T}\log_2 N(T)$$

where N(T) = Number of allowed signal sequences in a duration T.

An application of the channel capacity concept to an additive white Gaussian noise channel with *B* Hz bandwidth and signal-to-noise ratio *S/N* is the Shannon–Hartley theorem:

$$C = B\log\left(1+\frac{S}{N}\right)$$

Where C is measured in bits per second if the logarithm is taken in base 2, assuming B is in hertz; the signal and noise powers S and N are measured in watts or volts^2, so the signal-to-noise ratio here is expressed as a power ratio, *not* in decibels (dB); since figures are often cited in dB, a conversion may be needed. For example, 40 dB is a power ratio of $10^{40/10} = 10^4 = 10000$. And if For example, 40 dB is a power ratio of $10^{20/10} = 10^2 = 100$

8. a) What do you mean by News Value and Information Content?

[MODEL QUESTION]

Answer:

Information contained in a message is a separate quantity than its probability of occurrence. Information contained in a symbol should follow some properties.

1. Information 'I' should be always positive.
2. For a symbol with probability approaching its highest value 1, amount of information in it should approach in lowest value.
3. For two different symbols $x_i \; and \, x_j$ with respect probabilities P_i and P_j, the one with lower probability should contain more information i.e., $P_i < P_j$.

   The average amount of information in a message is called its entropy.
   So mathematically we can say that,

$$I = \log\left(\frac{1}{P}\right)$$

and $H = \sum_{i=1}^{n} P_i I_i$ where $I =$ Information

$P =$ Probability of occurence.

Unit of Information

The information content of a symbol (y_i), denoted by $I(y_i)$ is defined by

$$I(y_i) = \log_b \frac{1}{P(y_i)} = -\log_b P(y_i)$$

Where $P(y_i)$ is the probability of occurrence of symbol y_i. The unit of $I(y_i)$ is the bit (binary unit) if $b=2$. Hartley or decit if $b=10$ and nat (natural unit) if $b=e$. Here the unit bit (abbreviated 'b') is a measure of information content and is not be confused with the term 'bit' meaning binary digit.

b) Discuss about Error Control and Correcting Codes in coding theory. **[MODEL QUESTION]**

Answer:

The two important objectives of digital communication is to minimize errors and to maintain data security. The bit error rate (BER) in digital communication system depends on ratio of the signal energy per bit to noise power spectral density i.e., E_b/N_o. Increasing the ratio reduces the BER. But in actual practice there is a limit to the value of E_b/N_o. For a fixed value of E_b/N_o a modulation scheme may give unacceptably high BER. In such a case the best solution for reducing BER and hence improve error performance is to use what is known as error-control coding.

Error control coding aims at systematic addition of extra or redundant digits to the transmitted message. Though the addition of redundancy increases the transmission bandwidth it does the very important job of error detection and correction. The extra bits do not convey any information by themselves but make it possible to detect or correct errors in the received message. It may be recalled that the channel encoder in the transmitter accepts message bits and adds redundancy in a systematic manner. The channel decoder in the receiver utilizes this redundancy to make the correct decision. The channel encoder and the channel decoder together minimize the effect of noise in the digital communication system. However, apart from increasing the transmissions bandwidth, error-control coding also increases the system complexity.

The codes that control the errors in a digital communication system are called error-control codes. There are some codes, which can only detect errors without correction. These codes are called error detection codes. Other codes are capable of both error detection and correction. Such codes are called error correction codes. Coding for error detection, without correction, is simpler than error-correction coding. Accordingly there are two types of error control, namely, Automatic Repeat Request or ARQ and Forward Error Correction or FEC.