3.3. Information Sources

An information source may be viewed as an object which produces an event, the outcome of which is selected at random according to a probability distribution. A practical source in a communication system is a device which produces messages, and it can be either analog or discrete. In this chapter, we deal mainly with the discrete sources since analog sources can be transformed to discrete sources through the use of sampling and quantization techniques, described in chapter 10. As a matter of fact, a discrete information source is a source which has only a finite set of symbols as possible outputs. The set of source symbols is called the **source alphabet**, and the elements of the set are called **symbols** or **letters.**

Information sources can be classified as having memory or being memoryless. A source with memory is one for which a current symbol depends on the previous symbols. A memoryless source is one for which each symbol produced is independent of the previous symbols.

A discrete memoryless source (DMS) can be characterized by the list of the symbols, the probability assignment to these symbols, and the specification of the rate of generating these symbols by the source.

3.4. Information Content of a Discrete Memoryless Source (DMS)

The amount of information contained in an event is closely related to its uncertainty. Messages containing knowledge of high probability of occurrence convey relatively little information. We note that if an event is certain (that is, the event occurs with probability 1), it conveys zero information. Thus, a mathematical measure of information should be a function of the probability of the outcome and should satisfy the following axioms:

(i) Information should be proportional to the uncertainty of an outcome.

(ii) Information contained in independent outcomes should add.

3.5. Information Content of a Symbol (*i.e.*, Logarithmic Measure of Information)

Let us consider a discrete memoryless source (DMS) denoted by X and having alphabet $\{x_1, x_2, \ldots x_m\}$. The *information content* of a symbol x_i, denoted by $I(x_i)$ is defined by

$$I(x_i) = \log_b \frac{1}{P(x_i)} = -\log_b P(x_i) \qquad \ldots(3.1)$$

where $P(x_i)$ is the probability of occurrence of symbol x_i.

Note that $I(x_i)$ satisfies the following properties:

$$I(x_i) = 0 \text{ for } P(x_i) = 1 \qquad \ldots(3.2)$$

$$I(x_i) \geq 0 \qquad \ldots(3.3)$$

$$I(x_i) > I(x_j) \text{ if } P(x_i) < P(x_j) \qquad \ldots(3.4)$$

$$I(x_i, x_j) = I(x_i) + I(x_j) \text{ if } x_i \text{ and } x_j \text{ are indepedent} \qquad \ldots(3.5)$$

The unit of $I(x_i)$ is the bit (binary unit) if $b = 2$, Hartely or decit if $b = 10$, and nat (natural unit) if $b = e$. It is standard to use $b = 2$. Here the unit bit (abbreviated "b") is a measure of information content and is not to be confused with the term 'bit' meaning "binary digit". The conversion of these units to other units can be achieved by the following relationships.

$$\log_2 a = \frac{\ln a}{\ln 2} = \frac{\log a}{\log 2} \qquad \ldots(3.6)$$

Example 3.1. 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.

Solution: We know that the information content of each symbol is given as

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

Thus, we can write

$$I(x_1) = \log_2 \frac{1}{\frac{1}{2}} = \log_2 (2) = 1 \text{ bit}$$

$$I(x_2) = \log_2 \frac{1}{\frac{1}{4}} = \log_2 2^2 = 2 \text{ bits}$$

$$I(x_3) = \log_2 \frac{1}{\frac{1}{8}} = \log_2 2^3 = 3 \text{ bits}$$

$$I(x_4) = \log_2 \frac{1}{\frac{1}{8}} = 3 \text{ bits} \qquad \text{Ans.}$$

Example 3.2. Calculate the amount of information if it is given that $P(x_i) = \frac{1}{4}$.

Solution: We know that amount of information is given as,

$$I(x_i) = \log_2 \frac{1}{P(x_i)} = \frac{\log_{10} \frac{1}{P(x_i)}}{\log_{10} 2}$$

Substituting given value of $P(x_i)$ in above equation, we get

$$I(x_i) = \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits} \qquad \text{Ans.}$$

3.6. Entropy (*i.e.*, Average Information)

In a practical communication system, we usually transmit long sequences of symbols from an information source. Thus, we are more interested in the average information that a source produces than the information content of a single symbol.

In order to get the information content of the symbol, we take notice of the fact that the flow of information in a system can fluctuate widely because of randomness involved into the selection of the symbols. Thus, we require to talk about the average information content of the symbols in a long message.

Thus, for quantitative representation of average information per symbol we make the following assumptions:

(i) The source is stationary so that the probabilities may remain constant with time.

(ii) The successive symbols are statistically independent and come form the source at a average rate of r symbols per second.

The mean value of $I(x_i)$ over the alphabet of source X with m different symbols is given by

$$H(X) = E\,[I(x_i)] = \sum_{i=1}^{m} P(x_i)I(x_i)$$

or

$$H(X) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i) \text{ b/symbol} \quad \text{...(3.7)}$$

The quantity $H(X)$ is called the **entropy** of source X. It is a measure of the *average information content per source symbol*. The source entropy $H(X)$ can be considered as the average amount of uncertanity within source X that is resolved by use of the alphabet.

It may be noted that for a binary source X which generates independent symbols 0 and 1 with equal probability, the source entropy $H(X)$ is

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1\text{b/symbol} \quad \text{...(3.8)}$$

The source entropy $H(X)$ satisfies the following relation:

$$0 \le H(X) \le \log_2 m \quad \text{...(3.9)}$$

where m is the size (number of symbols) of the alphabet of source X. The lower bound corresponds to no uncertainty, which occurs when one symbol has probability $P(x_i) = 1$ while $P(x_j) = 0$ for $j \ne i$, so X emits the same symbol x_i all the time. The upper bound corresponds to the maximum uncertainty which occurs when $P(x_i) = 1/m$ for all i, that is, when all symbols are equality likely to be emitted by X.

3.7. Information Rate

If the time rate at which source X emits symbols is r (symbols s), the *information rate* R of the source is given by

$$R = rH(X) \quad b/s \qquad \text{...(3.10)}$$

Here R is information rate.

$H(X)$ is Entropy or average information

and r is rate at which symbols are generated.

Information rate R is represented in average number of bits of information per second. It is calculated as under:

$$R = \left[r \text{ in } \frac{\text{symbols}}{\text{second}}\right] \times \left[H(X) \text{ in } \frac{\text{Information bits}}{\text{symbol}}\right] \qquad \text{...(3.11)}$$

or

$$R = \text{Information bits/second} \qquad \text{...(3.12)}$$

or

Example 3.9. 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)

Solution: (i) We know that entropy is given by

$$H(X) = -\sum_{i=1}^{4} P(x_i)\log_2[P(x_i)]$$

Substituting values and simplifying, we get

$$H(X) = -0.4\log_2 0.4 - 0.3\log_2 0.3 - 0.2\log_2 0.2 - 0.1\log_2 0.1$$

Solving, we get

$$H(X) = 1.85 \text{ b/symbol} \qquad \textbf{Ans.}$$

(ii) Now, we have

$$P(x_1\,x_2\,x_1\,x_3) = (0.4)(0.3)(0.4)(0.2) = 0.0096$$

Therefore,

$$I(x_1x_2x_1x_3) = -\log_2 0.0096 = 6.70 \text{ b/symbol}^* \qquad \textbf{Ans.}$$

Thus, $I(x_1x_2x_1x_3) < 7.4\,[= 4H(X)]$ b/symbol

Also, $P(x_4x_3x_3x_2) = (0.1)(0.2)^2(0.3) = 0.0012$

Therefore

$$I(x_4x_3x_3x_2) = -\log_2 0.0012 = 9.70 \text{ b/symbol}$$

We conclude that

$$I(x_4x_3x_3x_2) > 7.4[= 4H(X)] \text{ b/symbol} \qquad \textbf{Ans.}$$

3.15. Channel Capacity *(U.P. Tech, Sem. Exam., 2005-2006)*

We know that the bandwidth and the noise power place a restriction upon the rate of information that can be transmitted by a channel. It may be shown that in a channel which is disturbed by a white Gaussian noise, one can transmit information at a rate of C bits per second, where C is the the channel capacity and is expressed as

$$C = B\log_2\left(1+\frac{S}{N}\right) \qquad ...(3.51)$$

In this expression,

B = channel bandwidth in Hz

S = Signal power

N = Noise power

3.19. The Source Coding

A conversion of the output of a DMS into a sequence of binary symbols (*i.e.*, binary code word) is called **source coding.** The device that performs this conversion is called the **source encoder** as shown in figure 3.15.

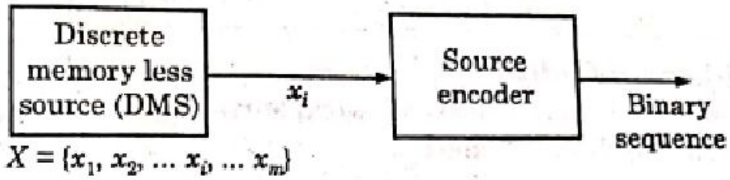


Fig. 3.15. *Block diagram for source Coding.*

An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source.

3.19.1. The Code Length and Code Efficiency

Let X be a DMS with finite entropy $H(X)$ and an alphabet $\{x_1, \ldots\, x_m\}$ with corresponding probabilities of occurrence $P(x_i)(i = 1,\ldots,m)$.

Let the binary codeword assigned to symbol x_i by the encoder have length n_i, measured in bits. The length of a codeword is the number of binary digits in the codeword. The average codeword length L, per source symbol is given by

$$L = \sum_{i=1}^{m} P(x_i)n_i \qquad \ldots(3.51)$$

The parameter L represents the average number of bits per source symbol used in the source coding process.

Also, the *code efficiency* η is defined as

$$\eta = \frac{L_{min}}{L} \qquad \ldots(3.52)$$

where L_{min} is the minimum possible value of L. When η approaches unity, the code is said to be *efficient*.

The code *redundancy* γ is defined as

$$\gamma = 1 - \eta \qquad \ldots(3.53)$$

3.19.2. The Source Coding Theorem

The source coding theorem states that for a DMS X, with entropy $H(X)$, the average codeword length L per symbol is bounded as

$$L \geq H(X) \qquad \text{...(3.54)}$$

and further, L can be made as close to $H(X)$ as desired for some suitably chosen code.

Thus, with $L_{min} = H(X)$, the code efficiency can be rewritten as

$$\eta = \frac{H(X)}{L} \qquad \text{...(3.55)}$$

3.20. Entropy Coding

The design of a variable-length code such that its average codeword length approaches the entropy of DMS is often referred to as **entropy coding.** In this section, we present two examples of entropy coding.

3.20.1. Shannon-Fano Coding

An efficient code can be obtained by the following simple procedure, known as **Shannon-Fano algorithm:**

1. List the source symbols in order of decreasing probability.
2. Partition the set into two sets that are as close to equiprobables as possible, and assign 0 to the upper set 1 to the lower set.
3. Continue this process, each time partitioning the sets with as nearly equal probabilities as possible until further partitioning is not possible.

An example of Shannon-Fano encoding is shown in Table 3.6. Note in Shannon-Fano encoding the ambiguity may arise in the choice of approximately equiprobable sets.

Table 3.6. Shannon-Fano Encoding

x_i	$P(x_i)$	*Step 1*	*Step 2*	*Step 3*	*Step 4*	*Code*
x_1	0.30	0	0			00
x_2	0.25	0	1			01
x_3	0.20	1	0			10
x_4	0.12	1	1	0		110
x_4	0.08	1	1	1	0	1110
x_6	0.05	1	1	1	1	1111

$$H(X) = 2.36 \text{ b/symbol}$$
$$L = 2.38 \text{ b/symbol}$$
$$\eta = H(X)/L = 0.99$$

3.20.2. The Huffman Encoding (*U.P. Tech., Sem. Examination, 2003-04*)

In general, Huffman encoding results in an optimum code. Thus, it is the code that has the highest efficiency. The Huffman encoding procedure is as follows:

1. List the source symbols in order of decreasing probability.
2. Combine the probabilities of the two symbols having the lowest probabilities, and reorder the resultant probabilities, this step is called reduction 1. The same procedure is repeated until there are two ordered probabilities remaining.
3. Start encoding with the last reduction, which consist of exactly two ordered probabilities. Assign 0 as the first digit in the codewords for all the source

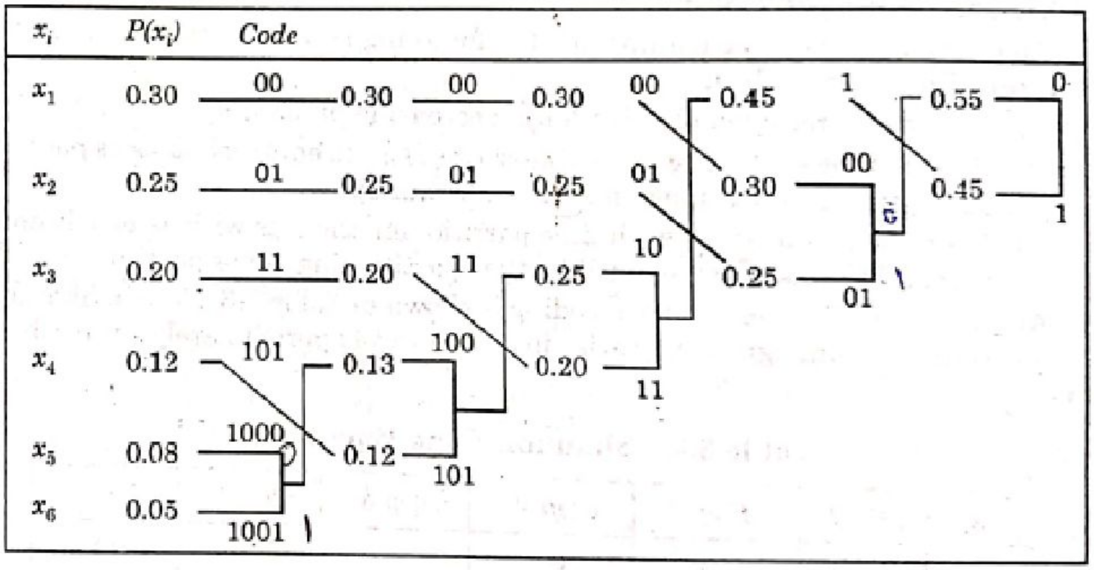
symbols associated with the first probability; assign 1 to the second probability.

4. Now go back and assign 0 and 1 to the second digit for the two probabilities that were combined in the previous reduction step, retaining all assignments made in Step 3.
5. Keep regressing this way until the first column is reached.

An example of Huffman encoding is shown in Table 3.7

$$H(X) = 2.36 \text{ b/symbol}$$
$$L = 2.38 \text{ b/symbol}$$
$$\eta = 0.99$$

Table 3.7. Huffman Encoding



Example 3.44. A DMS X has four symbols x_1, x_2, x_3, and x_4 with $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, and $P(x_3) = P(x_4) = \frac{1}{8}$. Construct a Shannon-Fano code for X; show that this code has the optimum property that $n_i = I(x_i)$ and that the code efficiency is 100 percent.

Solution: The Shannon-Fano code is constructed as follows (see Table 3.8).

$$I(x_3) = -\log_2 \frac{1}{8} = 3 = n_3$$

$$I(x_4) = -\log_2 \frac{1}{8} = 3 = n_4$$

We know that,

$$H(X) = \sum_{i=1}^{4} P(x_i) I(x_i)$$

or

$$H(X) = \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1.75$$

$$L = \sum_{i=1}^{4} P(x_i)\, n_i = \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1.75$$

Also

$$\eta = \frac{H(X)}{L} = 1 = 100\% \quad \textbf{Ans.}$$

Example 3.45. A DMS X has five equally likely symbols.

(*i*) Construct a Shannon-Fano code for *X*, and calculate the efficiency of the code.

(*ii*) Construct another Shannon-Fano code and compare the results.

(*iii*) Repeat for the Huffman code and compare the results.

(Madras University, 1998)

Solution: A Shannon-Fano code [by choosing two approximately equiprobable (0.4 versus 0.6) sets] is constructed as follows (see Table 3.9)

Table 3.9.

x_i	$P(x_i)$	*Step 1*	*Step 2*	*Step 3*	*Code*
x_1	0.2	0	0		00
x_2	0.2	0	1		01
x_3	0.2	1	0		10
x_4	0.2	1	1	0	110
x_5	0.2	1	1	1	111

$$H(X) = \sum_{i=1}^{5} P(x_i)\log_2 P(x_i) = 5(-0.2\log_2 0.2) = 2.32$$

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+2+2+3+3)\ 2.4$$

The efficiency η is $\quad \eta = \dfrac{H(X)}{L} = \dfrac{2.32}{2.4} = 0.967 = 96.7\%$ **Ans.**

(*ii*) Another Shannon-Fano code [by choosing another two approximately equiprobable (0.6 versus 0.4) sets] is constructed as follows (see Table 3.10)

11

Example 3.46 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 $P(x_5) = 0.1$.

(*i*) Construct a Shannon-Fano code for X, and calculate the efficiency of the code.

(*ii*) Repeat for the Huffman code and compare the results.

Solution: The Shannon-Fano code is constructed as follows (see Table 3.12)

$$H(X) = \sum_{i=1}^{5} P(x_i)n_i = 0.4(1) + 0.19(2) + 0.16(2) + 0.15(3) + 0.1(3) = 2.25$$

Also, $\quad \eta = \dfrac{H(X)}{L} = \dfrac{2.15}{2.25} = 0.956 = 95.6\%$ **Ans.**

x_i	$P(x_i)$	*Step 1*	*Step 2*	*Step 3*	*Code*
x_1	0.2	0	0		00
x_2	0.2	0	1	0	010
x_3	0.2	0	1	1	011
x_4	0.2	1	0		10
x_5	0.2	1	1		11

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

Since the average codeword length is the same as that for the code of part (*a*), the efficiency is the same. **Ans.**

(*iii*) The Huffman code is constructed as follows (see Table 3.11)

$$L = \sum_{i=1}^{5} P(x_i)n_i = 0.2(2+3+3+2+2) = 2.4$$

Since the average code word length is the same as that for the Shannon-Fano code, the efficiency is also the same.

Table 3.11.

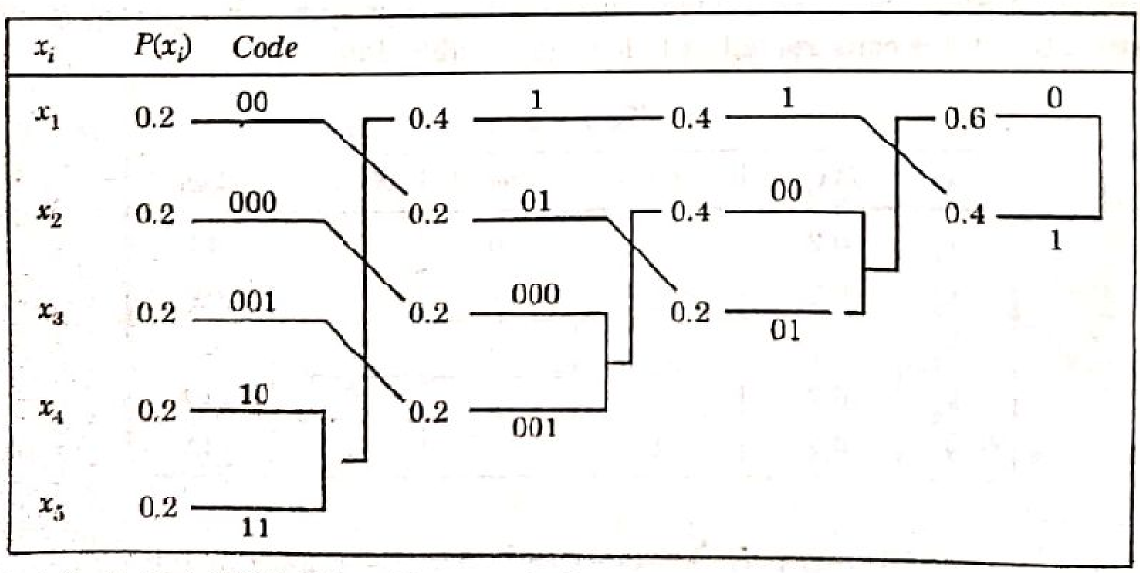


Table 3.12.

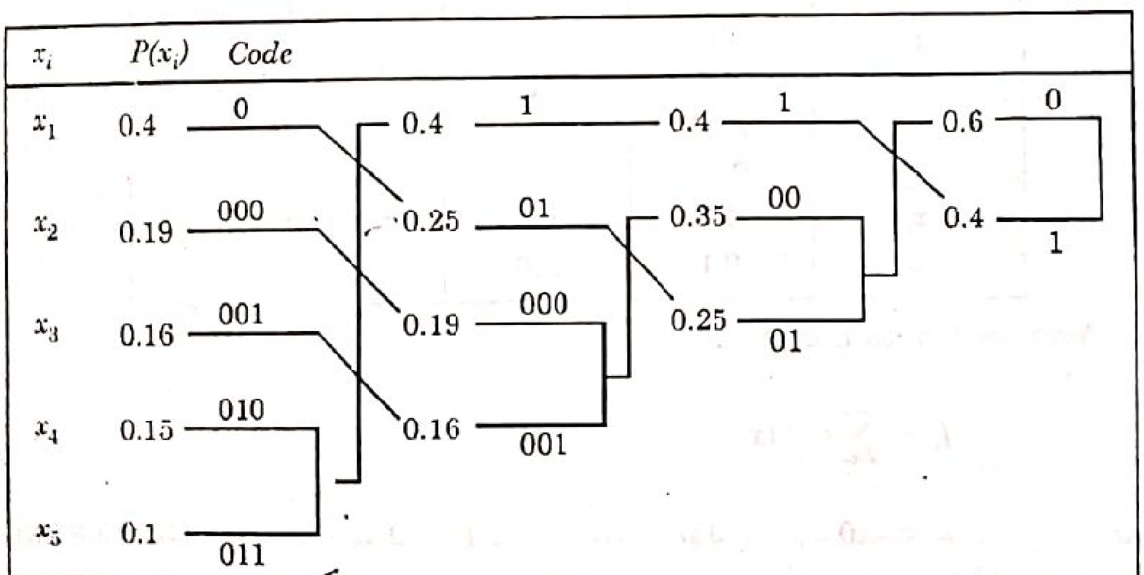
x_i	$P(x_i)$	*Step 1*	*Step 2*	*Step 3*	*Code*
x_1	0.4	0	0		00
x_2	0.19	0	1		01
x_3	0.16	1	0		10
x_4	0.15	1	1	0	110
x_5	0.1	1	1	1	111

(*ii*) The Huffman code is constructed as follows (see Table 3.13):

$$L = \sum_{i=1}^{5} P(x_i)n_i$$

or
$$L = 0.4(1) + (0.19 + 0.16 + 0.15 + 0.1)(3) = 2.2 \quad \textbf{Ans.}$$

Table 3.13.



or
$$\eta = \frac{H(X)}{L} = \frac{2.15}{2.2}$$

or
$$\eta = 0.977 = 915.7\%$$

The average codeword length of the Huffman code is shorter than that of the Shannon-Fanco code, and thus the efficiency is higher than that of the Shannon-Fano code. **Ans.**

Fano code. Ans.

Example 3.47. Determine the Huffman code for the following messages with their probabilities given

x_1	x_2	x_3	x_4	x_5	x_6	x_7
0.05	0.15	0.2	0.05	0.15	0.3	0.1

Solution: Arranging and groping of messages is done as shown below:

x_i	$p(x_i)$	Code										
x_6	0.3	00	0.3	CO	0.3	00	0.3	00	0.4	1	0.6	0
x_3	0.2	1.0	0.2	10	0.2	10	0.3	01	0.3	00	0.4	1
x_2	0.15	101	0.15	010	0.2	11	0.2	10	0.3	01		
x_5	0.15	011	0.15	011	0.15	010	0.2	11				
x_7	0.1	110	0.1	110	0.15	011						
x_1	0.05	1100	0.1	111								
x_4	0.05	1111										

Message	*Prob.*	*Code*	*No. of bits in code*
x_1	0.05	1110	4
x_2	0.15	010	3
x_3	0.2	10	2
x_4	0.05	1111	4
x_5	0.15	011	3
x_6	0.3	00	2
x_7	0.1	110	3

$\therefore$ Average length L is given by

$$L = \sum_{i=1}^{7} n_i P(x_i)$$

or $$L = 4(0.05 + 0.05) + 3(0.15 + 0.15 + 0.1) + 2(0.2 + 0.3) = 2.6 \text{ bits}$$

Entropy $H(X)$ is given by

$$H(X) = \sum_{i=1}^{7} P(x_i) \log_2 \frac{1}{P(x_i)}$$

$$= 0.3 \log_2 \left(\frac{1}{0.3}\right) + 0.2 \log_2 \left(\frac{1}{0.2}\right) + 0.3 \log \left(\frac{1}{0.15}\right) + 0.1 \log_2 \left(\frac{1}{0.3}\right) + 0.1 \log_2 \left(\frac{1}{0.05}\right)$$

$$= 2.57 \text{ bits}$$

$$\eta = \frac{H(X)}{L \log_2 M} = \frac{2.57}{2.6 \log_2 2} = \frac{2.57}{2.6} = 98.85\%$$