Therefore we may write

$$H(X) = -\sum_{i=1}^{N} p\{x_i\} \log p\{x_i\} < -\sum_{i=1}^{N} p\{x_i\} \log \frac{D^{-n_i}}{\sum_{k=1}^{N} D^{-n_k}}$$

$$= \log \sum_{k=1}^{N} D^{-n_k} - \sum_{i=1}^{N} p\{x_i\} \log D^{-n_i} \tag{1.142}$$

or

$$H(X) \leq \log \sum_{k=1}^{N} D^{-n_k} + \sum_{i=1}^{N} p\{x_i\} n_i \log D$$

Applying the theorem of section 1.9 yields

$$H(X) \leq \log 1 + \log D \sum_{i=1}^{N} p\{x_i\} n_i$$

$$\bar{L} \geq \frac{H(X)}{\log D} \tag{1.143}$$

This result is rather interesting as it clearly shows that even in the absence of noise no uniquely decipherable encoding procedure can be divised such that its average message length is less than a fixed number that is the ratio of the source entropy and logD. Here log D is the maximum possible entropy associated with the selected alphabet containing D characters or the capacity of the coding alphabet subject to the above constraint. This lower bound is not generally achieved unless the n_j are all approximately chosen integers.

In such a case we obtain an optimum encoding procedure, that is,

$$L_o = \frac{H(X)}{\log D} = \frac{\text{source entropy}}{\text{capacity of coding alphabet}} \tag{1.144}$$

While L_o, the lower of $\bar{L}$, may not always be reached by an encoding procedure, it is always possible to give an obtainable bound for $\bar{L}$. In fact, if we let

$$-\frac{\log p\{x_k\}}{\log D} \leq n_k < -\frac{\log p\{x_k\}}{\log D} + 1, \quad k = 1, 2, \ldots N, \quad D \geq 2$$

we rest assured that

$$\frac{H(X)}{\log D} \le \bar{L} = \sum_{k=1}^{N} p\{x_k\}.n_i < 1 + \frac{H(X)}{\log D} \tag{1.145}$$

This relationship is called the noiseless coding theorem. It gives the upper and lower bounds on the average length of a uniquely decipherable code that can be constructed for a source with entropy H(X). It has already been noted that any source with symbol probabilities satisfying equation

$$P(x_i) = D^{n_i}, \quad i = 1, 2, \ldots N \tag{1.146}$$

can be encoded into a code of absolutely minimum average length, H(x) / log D. Equation (1.145) guarantees that for an arbitrary source, the average code-word length can be constrained no more than one code element longer than this absolute minimum.

1.12 Shannon's Binary Coding

Shannon has suggested a binary encoding procedure based on equation(1.145). First we must reassure ourselves whether such codes exist. For this we employ the noiseless coding theorem. Equation(1.145) yields

$$2^{-n_k} \le p\{x_k\}, \; k = 1, 2, \ldots N \tag{1.147}$$

Thus we are sure that the desired code exists. Note that such codes will have the interesting property that their average length is constrained by

$$H(X) \le \bar{L} < H(X) + 1 \tag{1.148}$$

The following steps describe the method :

i) Write down the message ensemble in the order of non-increasing probability, say,

$$\{x_1, x_2, \ldots, x_N\}$$
$$p\{x_1\} \ge p\{x_2\} \ge \ldots \ge p\{x_N\} \tag{1.149}$$

ii) Compute the sequences

$$\alpha_1 = 0$$
$$\alpha_2 = p\{x_1\}$$
$$\alpha_3 = p\{x_2\} + p\{x_1\} = p\{x_2\} + \alpha_2$$
$$\alpha_4 = p\{x_3\} + p\{x_2\} + p\{x_1\} = p\{x_3\} + \alpha_3 \tag{1.150}$$

iii) Determine the set of integers which is the smallest integer's solution of the inequalities

$$2^{n_i} p\{x_i\} \geq 1, \ i = 1, 2, \ldots \quad (1.151)$$

iv) Expand the decimal numbers α_i in binary form to n_i places : that is, neglect the expansion beyond the n_i digits.

1.13 Construction of optimal codes

(a) *Shannon - fano Encoding*

This method of encoding is directed toward constructing reasonably efficient separable binary code for sources without memory. Let $\{X\}$ be the ensemble of the messages to be transmitted and [P] their corresponding probabilities:

$$[X] = [x_1, x_2, \ldots, x_n]$$

$$[P] = [p_1, p_2, \ldots, p_n]$$

It is desired to associate a sequence C_K of binary numbers of unspecified length n_k to each message x_k such that :

(i) No sequences of employed binary numbers C_k can be obtained from each other by adding more binary terms to the shorter sequence (prefix property)

(ii) The transmission of the encoded message is reasonably efficient, that is, 1 and 0 appear independently and with almost equal probabilities.

The Shannon - Fano encoding procedure is as follows:

The messages are first written in order of nonincreasing probabilities. Then the message set is partitioned into two most equiprobable subsets $\{x_1\}$ and $\{x_2\}$. A 0 is assigned to each message contained in one subset and a 1 to each of the remaining messages. The same procedure is repeated for subsets of $\{x_1\}$ and $\{x_2\}$. That is , $\{x_1\}$ will be partitioned into the subsets $\{x_{11}\}$ and $\{x_{12}\}$. Now the code word corresponding to message in x_{11} will start with 00 and that corresponding to a message in x_{12} will begin with 01. This procedure is continued until each subset contains only one message. Note that each digit 1 or 0 in each partitioning of the probability space appears with more or less equal probability, independent of the previous or subsequent partitioning; therefore the second requirement is also fulfilled.

This encoding procedure is said to be an optimum procedure minimizing the average length of messages. No other encoding procedure satisfying the above requirements can be found that leads to a smaller average number of digits per encoded message.

b) Huffman's Minimum Redundancy code

The procedure for code alphabet of arbitrary size D is described as follows:

(i) Arrange the M source symbols in order of decreasing probability.

(ii) Arrange the code elements in an arbitrary but fixed order i.e., a_1, a_2, . . ., a_D.

(iii) Sum the probabilities of the D least likely symbols and reorder the resulting M–(D–1) probabilities: this step is called reduction 1.

Note: For the binary case (D = 2), it will always be possible to accomplish this reduction in M-2 steps. When the size of the code alphabet is arbitrary, the last reduction will occur in exactly D ordered probabilities if, and only if,

$$M = D + n(D - 1)$$

where n is an integer. In case this relationship is not satisfied, the proper procedure to follow is to add dummy source symbols with zero probability. The entire encoding procedure is followed as before and , at the end, the dummy symbols are thrown away.

(iv) Start the encoding with the last reduction which consists of exactly D ordered probabilities: assign the element a_1 as the first digit in the code words for all the source symbols associated with the first probability, assign a_2 to the second probability and a_i to the ith probability.

(v) Proceed to the next - to - the - last reduction ; this reduction consists of D + (D – 1) ordered probabilities for a net gain of D – 1 probabilities. For the D new probabilities the first code digit has already been assigned and is the same for all of these D probabilities, assign a_1 as the second digit for all source symbols associated with the first of these D new probabilities ; assign a_2 as the second digit for the second of these D new probabilities and so on.

(vi) The encoding procedure terminates after 1 + n (D – 1) steps which is one more than the number of reductions.

Huffman has suggested a simple method for constructing separable codes with minimum redundancy for a set of discrete messages. Let [X] be the message ensemble, [P] the corresponding probability matrix, [D] the encoding alphabet, and L $[x_K]$ the length of the encoded message x_K.

Then

$$\bar{L} = E[L(x_k)] = \sum_{k=1}^{N} P\{x_k\}.L(x_k) \tag{1.152}$$

A minimum redundancy or an optimum code is one that leads to the lowest possible value of $\bar{L}$ for a given D. That is, distinct message must be encoded in uniquely decipherable words with the prefix property. To comply with these requirements, Huffman derives the following results :

a) For an optimum encoding, the longer code word should correspond to a message with lower probability; thus if for convenience the messages are numbered in order of nonincreasing probability,

$$P\{x_1\} \geq P\{x_2\} \geq P\{x_3\} \geq \ldots \geq P\{x_N\} \tag{1.153}$$

then $$L(x_1) \leq L(x_2) \leq L(x_3) \leq \ldots \leq L(x_N) \tag{1.154}$$

Indeed, if equation (1.154) is not met for two messages x_k and x_j, one may interchange their corresponding codes and arrive at a lower value of $\bar{L}$. Thus such codes cannot be of the optimum type.

b) For an optimum Code it is necessary that

$$L(x_{N-1}) = L(x_N) \tag{1.155}$$

For an optimum encoding, n_0, the number of least probable messages should be encoded in words of equal length, is the integer satisfying the requirements

$$\frac{N - n_0}{D - 1} = \text{integer} \quad 2 \leq n_o \leq D$$

c) Each sequence of length $L(x_N) - 1$ digits either must be used as an encoded word, or must have one of its prefixes used as an encoded word.

1.14 Applications of Huffmann´s coding

The Huffmann code discussed in section 1.13 is a prefix-free variable-length code which can achieve the shortest average code length $\bar{L}$ for a given input alphabet. The shortest average code length for a particular alphabet may be significantly greater than the entropy of the source alphabet. The inability to exploit the promised data compression is related to the source alphabet, not to the coding technique. Often the alphabet can be modified to form an extension code, and the same coding technique is then reapplied to achieve better compression performance. Compression performance is measured by *Compression ratio*. This measure is defined to be equal to the ratio of the average number of bits per sample before compression to the average number of bits per sample after compression. Extension codes offer a very powerful technique to include the effects of non-independent symbol sets. For example, in English text, adjacent leeters are highly correlated, very common pairs include

th	re	in
sh	he	e-
de	ed	s-
ng	at	r-
te	es	d-

where the dash represent a space. Similarly common English three tuples include

the	and	for
ing	ion	ess

Thus rather than perform Huffman coding on the individual letters, it is more efficient to extend the alphabet to include all 1-tuples plus common 2-tuples and 3-tuples, and then perform the coding on the extension code.

Run length codes:

In many applications, a sequence of symbols to be transmitted or stored is characterized by lengthy runs of specific symbols. Rather than code each symbol of a lengthy run, it makes sense to describe the run with an efficient substitution code. As an example, runs of spaces (the most common symbol in text) are encoded in many communication protocols by a control character followed by the character count. The IBM 3780 BISYNC protocol has an option to replace runs of spaces with a "GS" character for ASCII or "IGS" character for EBCDIC, followed by a count of 2 to 63. Longer runs are partitioned into successive runs of 63 characters. The run-length substitution coding can be applied to the original symbol alphabet or the binary representation of that alphabet. Run-length coding is particularly attractive for binary alphabets derived from specific sources. The most important commercial example is facsimile coding used for transmitting documents by instant electronic mail.

Huffman coding for Facsimile Transmission

Facsimile transmission is the process of transmitting a two-dimensional image as a sequence of successive line scans. The most common images are, documents containing text and figures. The postion of the scan lines and the position along a scan line are quantized into spatial locations that define a two-dimensional grid of picture elements called "Pixels". The standard CCITT document is defined to be of width 21cm × 29cm, alomst 8.5 in by 11 in. The spatial quantization for normal resolution is 1188 pixels/line and 1728 lines/document. The standard also defines a high-resolution quantisation of 2376 pixels/line with the same 1728 lines/ document. The total number of individual pixels for a normal-resolution facsimile transmission is 2,052,864 and is doubled for high resolution. For comparison, the number of pixels in NTSC standard commercial Television is 480 × 640 or 307,200. Thus facsimile has 6.7 or 13.4 times the resolution of a standard TV image.

The relative brightness or darkness of the scanned image at each position in the scan is quantized into two levels: B for black and W for white. Thus the signal observed during a scan line is a two-level pattern representing the B and W image intensity under scan. It is easy to see that a horizontal scan line across this sheet of paper will exhibit a pattern consisting of long runs of B and W levels. The standard CCITT run-length coding scheme to compress the run of B and W levels is based on a modified variable-length Huffman code. Two types of patterns are identified, runs of W and runs of B. Each run length is described by a "partitioned codeword". The first partition, called the "makup codeword" or most significant bits (MSB), identifies runs with lengths that are multiples of 64. The second partition, called the "terminating codeword" or least significant bits (LSB), identifies the length of the remaining run. Each run of B (or W) of length 0 through 63 is assigned a unique Huffmann code word, as is each run of length 64×K, K = 1,2,.......,27. A unique END OF LINE (EOL) is also defined in the code, which indicates that no black pixels follow, hence the next line should be started; this is akin to a carriage return on a typewriter.

Problems

Example 1.1

The inhabitants of a certain village are divided into two groups A and B. Half the people in group A always tell the truth, three-tenths always lie, *and two -tenths always refuse to answer. In group B, three - tenths of the people are truthful, half are liars, and two -tenths always refuse to answer. Let p be the probability that a person selected at random will belong to group A. Let I = I(p) be the information conveyed about a person's truth-telling status by specifying his group membership. Find the maximum possible value of I and the percentage of people in group A for which maximum occurs.*

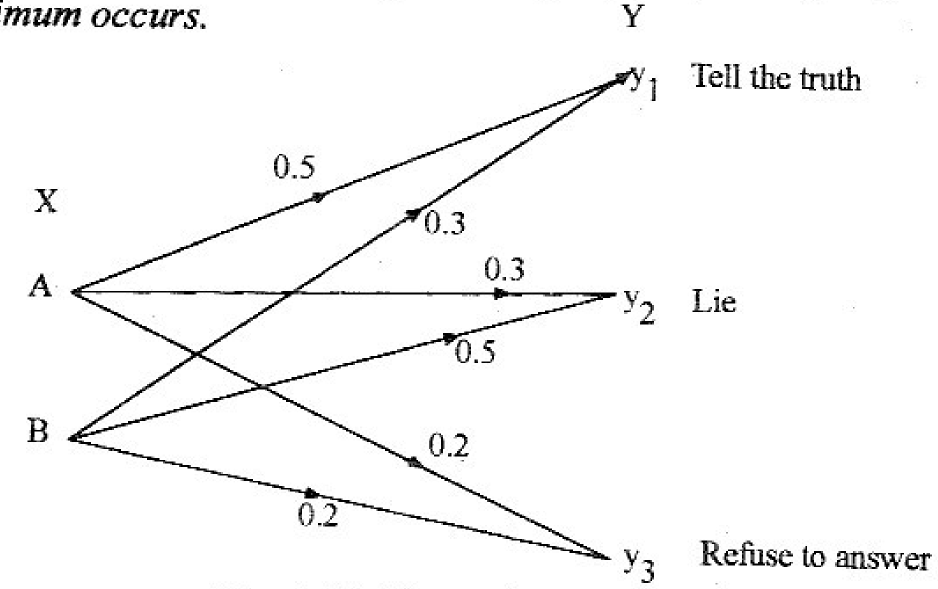


Fig. 1.12 Illustration of Probabilities

Solution

$$p(A) + p(B) = 1$$

$$p(y_1) = 0.5\, p(A) + 0.3[1 - p(A)]$$

$$p(y_2) = 0.3\, p(A) + 0.5[1 - p(A)]$$

$$p(y_3) = 0.2\, p(A) + 0.2[1 - p(A)] = 0.2$$

$$H(Y) = -\sum_{i=1}^{3} p(y_i).\log p(y_i)$$

$$= -[0.3 + 0.2p(A)]\log[0.3 + 0.2\, p(A)] - [0.5 - 0.2p(A)]\log[0.5 - 0.2\, p(A)] - 0.2\log 0.2$$

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

To maximize I(X/Y), differentiate with respect to p(A) and set the result equal to zero.

The maximum occurs when p(A) = 0.5.

$$I_{max} = -0.8\log 0.4 + 0.5\log 0.5 + 0.3\log 0.3$$
$$= 0.036.$$

Example 1.2

A single unbiased die is tossed once. If the face of the die is 1, 2, 3 or 4 an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information conveyed about the face of the die by the number of heads obtained.

Solution

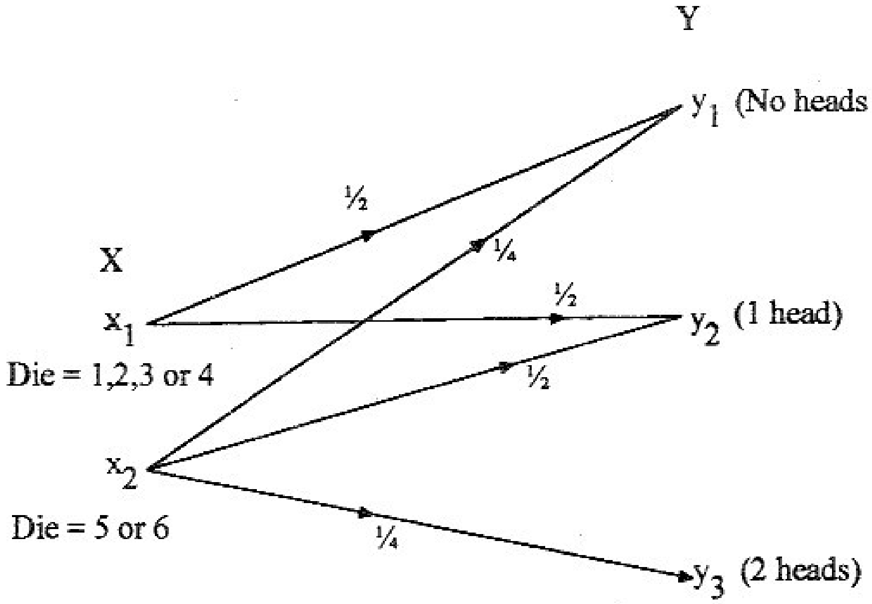


Fig. 1.13 Illustration of problem 1.2

$$I(X, Y) = H(Y) - H(Y/X) = H\left(\frac{5}{12}, \frac{1}{2}, \frac{1}{2}\right) - \frac{2}{3}H\left(\frac{1}{2}, \frac{1}{2}\right) - \frac{1}{3}H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$$

$$= -\frac{5}{12}\log\frac{5}{12} - \frac{1}{12}\log\frac{1}{12} - \frac{2}{3}$$

$$= 0.158$$

Example 1.3

Suppose that in a certain city, 3/4 of the high- school students pass and 1/4 fail. Of those who pass, 10 percent own cars, while 50 percent of the failing students own cars. All of the car-owning students belong to fraternities, while 40 percent of those who do not own cars but pass, as well as 40 percent of those who do not own cars but fail, belong to fraternities.

(a) How much information is conveyed about a student's academic standing by specifying whether or not he owns a car?

(b) How much information is conveyed about a student's academic standing by specifying whether or not he belongs to a fraternity?

(c) *If a student's academic standing, car - owning status, and fraternity status are transmitted by three successive binary digits, how much information is conveyed by each digit?*

Solution

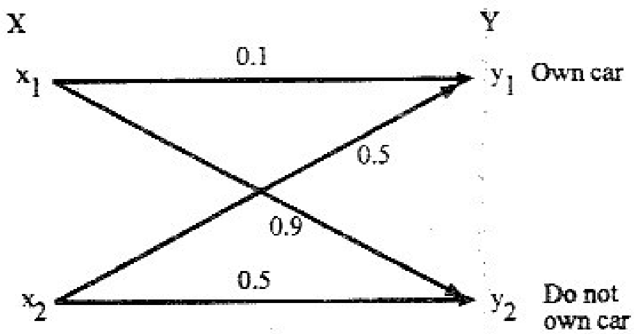


Fig. 1.14

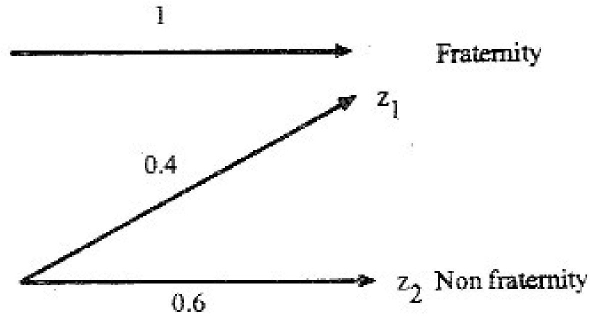


Fig. 1.15

$p(x_1) = 0.75 \quad p(x_2) = 0.25 \quad p(z_1/x_1) = 0.46 \quad p(z_2/x_1) = 0.54$
$p(y_1) = 0.2 \quad p(y_2) = 0.8 \quad p(z_1/x_2) = 0.7 \quad p(z_2/x_2) = 0.3$
$p(z_1) = 0.52 \quad p(z_2) = 0.48$

(a) $I(X, Y) = H(Y) - H(Y/X)$
$= H(0.2, 0.8) - 0.75\, H(0.1, 0.9) - 0.25\, H(0.5, 0.5)$
$= 0.12$

(b) $I(X, Z) = H(Z) - H(Z/X)$
$= H(0.52, 0.48) - 0.75\, H(0.46.\ 0.54) - 0.25\, H(0.7, 0.3)$
$= 0.03$

(c) First digit conveys $H(X) = H(0.75, 0.25) = 0.811$
Second digit conveys $H(Y/X) = 0.75\, H(0.1, 0.9) + 0.25\, H(0.5,0.5)$
$= 0.602$

Third digit conveys $H(Y/X, Y) = H(Z/Y)$ since $[p(z_k/x_i, y_j)] = p(z_k, y_j)$
$= 0.2\, H(1) + 0.8\, H(0.4, 0.6)$
$= 0.8\, H(0.4, 0.6)$
$= 0.777$

Example 1.4

If $p_1, p_2, \ldots, p_M$ and $q_1, q_2, \ldots, q_M$ are arbitrary positive numbers with $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$, prove that $-\sum_{i=1}^{M} p_i \log p_i \le -\sum_{i=1}^{M} p_i \log q_i$, with equality if and only if $p_i = q_i$ for all i.

Solution

For convenience we use natural logarithms instead of logarithms to the base 2. Since $\log_2 x = \log_2 e . \log_e x$, the inequality to be proved is unaffected by this change. The logarithm is a convex function; in other words ln x always lies below its tangent.

By considering the tangent at $x = 1$, we obtain $\ln x \le x - 1$ with equality if and only if $x = 1$.

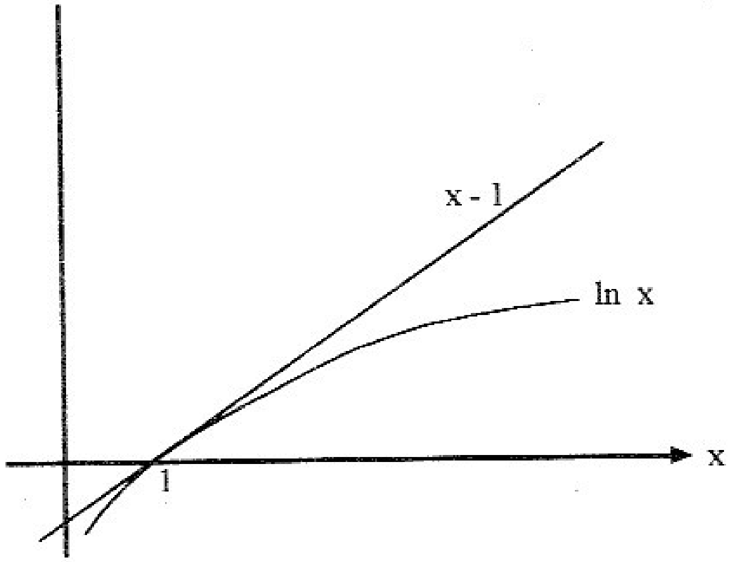


Fig. 1.16 Illustration of ln x

Thus $\ln\left(\frac{q_i}{p_i}\right) \le \left(\frac{q_i}{p_i}\right) - 1$ with equality if and only if $p_i = q_i$.

Multiplying the inequality by p_i and summing over i we obtain

$$\sum_{i=1}^{M} p_i \ln\frac{q_i}{p_i} \le \sum_{i=1}^{M} (q_i - p_i) = 1 - 1 = 0$$

with equality if and only if $p_i = q_i$ for all i.

Thus $$\sum_{i=1}^{M} p_i \ln q_i - \sum_{i=1}^{M} p_i \ln p_i \le 0$$

$$-\sum_{i=1}^{M} p_i \ln p_i \le -\sum_{i=1}^{M} p_i \ln q_i$$

Example 1.5

Prove that $H(p_1, p_2, \ldots, p_M) \leq \log M$ with equality if and only if all $p_i = 1/M$.

Solution

The application of the result of the previous problem (1.4) with all $q_i = 1/M$ yields

$$-\sum_{i=1}^{M} p_i \log p_i \leq -\sum_{i=1}^{M} p_i \log\frac{1}{M} = \log M \sum_{i=1}^{M} p_i = \log M$$

with equality if and only if $p_i = q_i = \dfrac{1}{M}$ for all i.

Example 1.6

Prove that H(X, Y) = H(X) + H(Y/X) = H(Y) + H(X/Y).

Solution:

$$H(X, Y) = -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j) \log p(x_i, y_j)$$

$$= -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j) \log p(x_i).p(y_j / x_i)$$

$$= -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j) \log p(x_i) - \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i, y_j) \log p(y_j / x_i)$$

$$= \sum_{i=1}^{M} p(x_i).\log p(x_i) + H(Y/X)$$

H (X, Y) = H(X) + H(Y/X)

Similarly, we can prove that H(X, Y) = H(Y) + H(X/Y)

A corresponding argument may be used to establish various identities involving more than two random variables.

$$H(X, Y, Z) = H(X) + H(Y/X) + H(Z/X, Y)$$
$$= H(X, Y) + H(Z/X, Y)$$
$$= H(X) + H(Y, Z/X).$$

$$H(X_1, \ldots X_n, Y_1, \ldots Y_n) = H(X_1, \ldots X_n) + H(Y_1, \ldots Y_m/X_1, \ldots X_n)$$

Example 1.7

Prove that $H(Y/X) \leq H(Y)$ with equality if and only if X and Y are independent.

Solution

We have $$H(X, Y) = H(X) + H(Y/X)$$

and $$H(X, Y) \leq H(X) + H(Y),$$

with equality if and only if X and Y are independent.

Therefore, $$H(X) + H(Y/X) \leq H(X) + H(Y)$$

$H(Y/X) \leq H(Y)$, with equality if and only if X and Y are independent.

Example 1.8

If X is a random variable which has the value 0 or 1 according as the unbiased or the two- headed coin is chosen and Y, the number of heads obtained in two tosses of the chosen coin, calculate the information conveyed about X by Y, given the diagram representing the experiment, together with various associated probability distributions (fig 1.17) Also prove that $I(X, Y) = I(Y, X)$

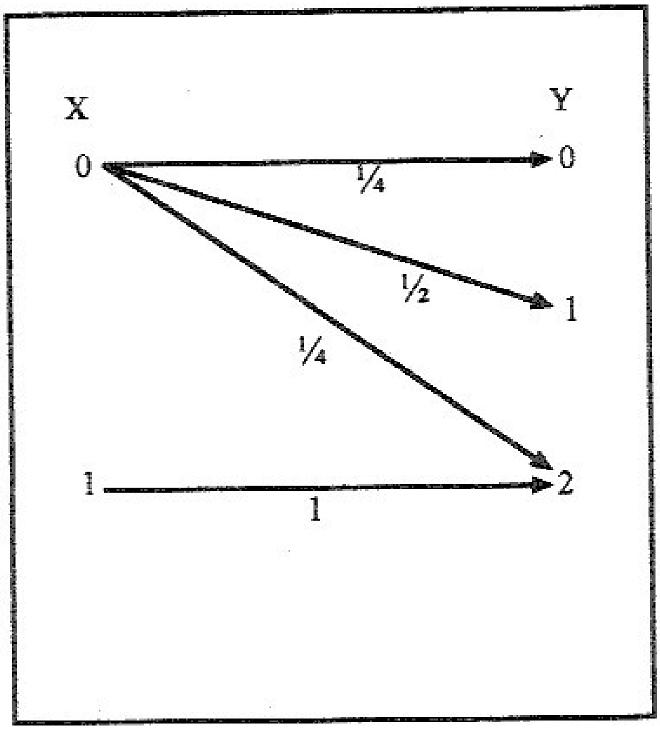


Fig. 1.17 A coin-tossing Experiment

$p(x = 0) = ½$	$p(x = 1) = ½$
$p(y = 0) = 1/8$	$p(x = 0/y = 0) = 1$
$p(y = 1) = 1/4$	$p(x = 0/y = 1) = 1$
$p(y = 2) = 5/8$	$p(x = 0/y = 2) = 1/5$

Solution

The initial uncertainity about the identity of the coin is H(X). After the number of heads is revealed, the uncertainity is H(X/Y).

Therefore the information conveyed about X by Y is

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

Here, $H(X) = \log 2 = 1$

$$H(X/Y) = p(y=0)\, H(x/y=0) + p(y=1)\, H(x/y=1) + p(y=2)\, H(x/y=2)$$

$$= \frac{1}{8}(0) + \frac{1}{4}(0) - \frac{5}{8}\left(\frac{1}{5}\log\frac{1}{5} + \frac{4}{5}\log\frac{4}{5}\right)$$

$$= 0.45$$

$$I(X,Y) = 0.55$$

Now, $H(Y) = -\frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{5}{8}\log\frac{5}{8} = 1.3$

$$H(Y/X) = p(X=0)\, H(Y/X=0) + p(X=1) H(Y/X=1)$$

$$= \frac{1}{2} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) + \frac{1}{2} H(1)$$

$$= 0.75$$

$$I(Y,X) = H(Y) - H(Y/X) = 0.55$$

Therefore, $I(X,Y) = I(Y,X)$

Example 1.9

Establish the following:

(a) $H(Y, Z/X) = H(Y/X) + H(Z/X)$

with equality if and only if $p(y_j, z_k / x_i) = p(y_j / x_i) \cdot p(z_k / x_i)$ *for all i, j, k.*

(b) $H(Y, Z/X) = H(Y/X) + H(Z/X, Y)$

(c) $H(Z/X,Y) \leq H(Z/X)$

Solution

(a) $H(Y/X) + H(Z/X)$

$$= \sum_i p(x_i, y_j, z_k) \log\left[p\left(y_j / x_i\right) + \log p\left(z_k / x_i\right)\right]$$

$$= -\sum_i p(x_i) \sum_{j,k} p(y_j, z_k / x_i) \log p(y_j / x_i) p(z_k / x_i)$$

From the result of problem (1.4), for each i,

$$-\sum_{j,k} p\left(y_j, z_k / x_i\right) \log p\left(y_j / x_i\right) p(z_k / x_i)$$

$$\geq -\sum_{j,k} p\left(\frac{y_j, z_k}{x_i}\right) \log p\left(\frac{y_j, z_k}{x_i}\right)$$

with equality iff

$$p\left(y_j, z_k / x_i\right) = p\left(y_j / x_i\right) p\left(z_k / x_i\right), \text{ for all j, k.}$$

Thus $H(Y/X) + H(Z/X) \geq H(Y,Z/X)$with equality if and only if $p(y_j, z_k / x_i) = p(y_j / x_i)\, p(z_k / x_i)$ for all i, j, k.

(b) $H(Y, Z / X) = -\sum_{i,j,k} p(x_i, y_j, z_k) \log p(y_j, z_k / x_i)$

$$= -\sum_{i,j,k} p(x_i, y_j, z_k) \log\left[p\left(y_j / x_i\right) p\left(z_k / x_i, y_j\right)\right]$$

$$= H(Y/X) + H(Z/X, Y)$$

(c) $H(Z / X, Y) = H(Y, Z / X) - H(Y/X)$ [by (b)]

$$\leq H(Y / Z) + H(Z / X) - H(Y / X) \qquad \text{[by (a)]}$$

So, $H(Z / X, Y) \leq H(Z/X)$ with equality if and only if

$$p\left(y_j, z_k / x_i\right) = p\left(y_j / x_i\right) \cdot p\left(z_k / x_i\right) \text{ for all i, j, k.}$$

Example 1.10

Given a discrete random variable X with values $x_1, \ldots, x_M$, define a random variable Y by $Y = g(X)$ where g is an arbitrary function. Show that $H(Y) \le H(X)$. Under what conditions on the function g will there be equality?

Solution

$$H(X, Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$$

But $H(Y/X) = 0$ since Y is determined by X.

Hence $$H(Y) \le H(X).$$

$H(Y) = H(X)$ if and only if $H(X/Y) = 0$; that is, if and only if X is determined by Y, that is, for each y_j there is an x_i such that $P(x_i/y_j) = 1$.

Hence $H(Y) = H(X)$ if and only if g is one to one on $\{x_1, \ldots x_M\}$, that is $x_i \ne x_j$ implies $g(x_i) \ne g(x_j)$.

Example 1.11

Show that if $h(p)$, $0 < p \le 1$, is a continuous function such that

$\sum_{i=1}^{M} p_i h(p_i) = -C \sum_{i=1}^{M} p_i \log p_i$, *for all M and all $p_1, \ldots, p_M$ such that $p_i > 0$,*

$\sum_{i=1}^{M} p_i = 1$, *then $h(p) = -C \log p$.*

Solution

Let $$h'(p) = h(p) + C \log p.$$

Then $$\sum_{i=1}^{M} p_i h'(p_i) = 0.$$

Taking all $p_i = \dfrac{1}{n}$, we have

$$h'\left(\frac{1}{n}\right) = 0, \quad n = 1, 2, \ldots$$

If r/s is a rational number, then, taking

$$p_1 = \frac{r}{s}; \quad p_2 = p_3 = \ldots = p_{s-r+1} = 1/s$$

We have $$\frac{r}{s}\, h'\left(\frac{r}{s}\right) + \frac{s-r}{s} h'\left(\frac{1}{s}\right) = 0$$

Hence $$h'\left(\frac{r}{s}\right) = 0.$$

The result now follows by continuity.

Example 1.12

Given a function h(p), $0 < p \le 1$, satisfying

(a) $h(p_1\, p_2) = h(p_1) + h(p_2)$, $0 < p_1 \le 1, 0 < p_2 \le 1$.

(b) h(p) is a monotonically decreasing and continuous function of p, $0 < p \le 1$.

Show that the only function satisfying the given conditions is $h(p) = -C \log_b p$ where $C > 0$, $b > 1$.

Solution

Let $$f(n) = h\left(\frac{1}{n}\right).$$

Then the requirements (a) and (b) imply $f(nm) = f(n) + f(m)$, and $n < m \Rightarrow f(n) < f(m)$, $n, m = 1, 2, \ldots$.

We have the theorem,

The only function satisfying the axioms of uncertainity measure is

$$H(p_1, \ldots p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

Hence $$f(n) = C \log_b n, \text{ where } C > 0, b > 1.$$

If p is a rational number $\frac{r}{s}$, then

$$h\left(\frac{1}{s}\right) = h\left(\frac{r}{s}.\frac{1}{r}\right) = h\left(\frac{r}{s}\right) + h\left(\frac{1}{r}\right)$$

Thus $$h\left(\frac{r}{s}\right) = f(s) - f(r) = -C \log\left(\frac{r}{s}\right)$$

Hence h(p) = – C log p for rational p.

The general assertion follows by continuity.

Example 1.13

Given below is the noise characteristic. Determine the rate of transmission through this channel.

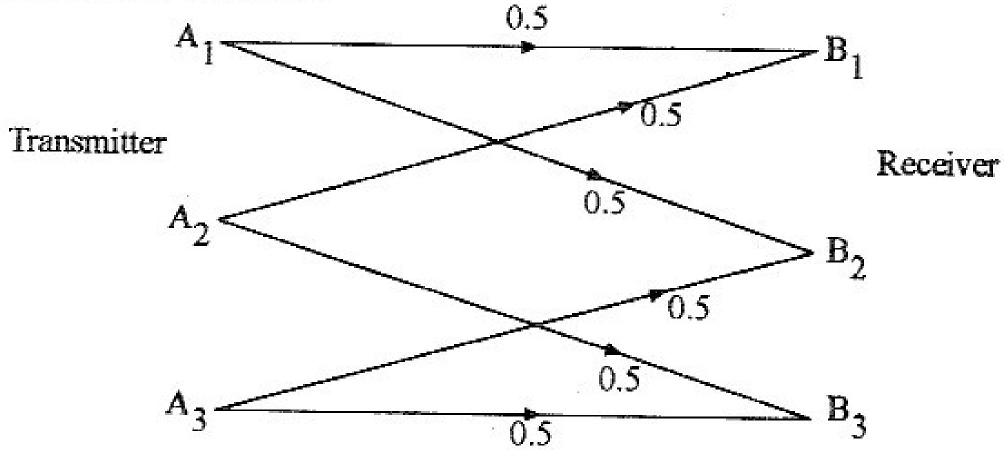


Fig. 1.18 Communication channel for problem 1.13

$$p(A_1) = 0.6\ ;\ p(A_2) = 0.3\ ;\ p(A_3) = 0.1$$

Solution

$$P(Y/X) = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix} \end{array}$$

Joint Matrix

$$p(X, Y) = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} (0.5)(0.6) & (0.5)(0.6) & 0 \\ (0.5)(0.3) & 0 & (0.5)(0.3) \\ 0 & (0.5)(0.1) & (0.5)(0.1) \end{pmatrix} \end{array}$$

$$= \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 0.30 & 0.30 & 0 \\ 0.15 & 0 & 0.15 \\ 0 & 0.05 & 0.05 \end{pmatrix} \end{array}$$

From p(X, Y) we obtain

$$p(B_1) = 0.45,\ p(B_2) = 0.35,\ p(B_3) = 0.20$$

$$\begin{aligned} H(Y) &= -\sum_{i=1}^{n} p(y_j) \log p(y_j) \\ &= -0.45 \log 0.45 - 0.35 \log 0.35 - 0.20 \log 0.20 \\ &= 0.5184 + 0.5301 + 0.4643 \\ H(Y) &= 1.5128 \text{ bits/symbol} \end{aligned}$$

Note that if each of the transmitted symbols is considered to be a message then the units of H(Y) will become bits/message.

$$\begin{aligned} H(Y/X) &= -\sum_{j=1}^{n}\sum_{i=1}^{m} p(x_i, y_j) \log p\left(y_j / x_i\right) \\ &= -0.30 \log 0.5 - 0.3 \log 0.5 - 0.15 \log 0.5 \\ &\quad - 0.15 \log 0.5 - 0.05 \log 0.5 - 0.05 \log 0.5 \\ &= -1 \log 0.5 \\ &= 1 \text{ bit/symbol} \end{aligned}$$

Rate of Transmission is given by

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y/X) \\ &= 1.5128 - 1 \\ I(X, Y) &= 0.5128 \text{ bits/symbol.} \end{aligned}$$

Example 1.14

Given the noise characteristic of a communication channel, evaluate the rate of transmission through the channel.

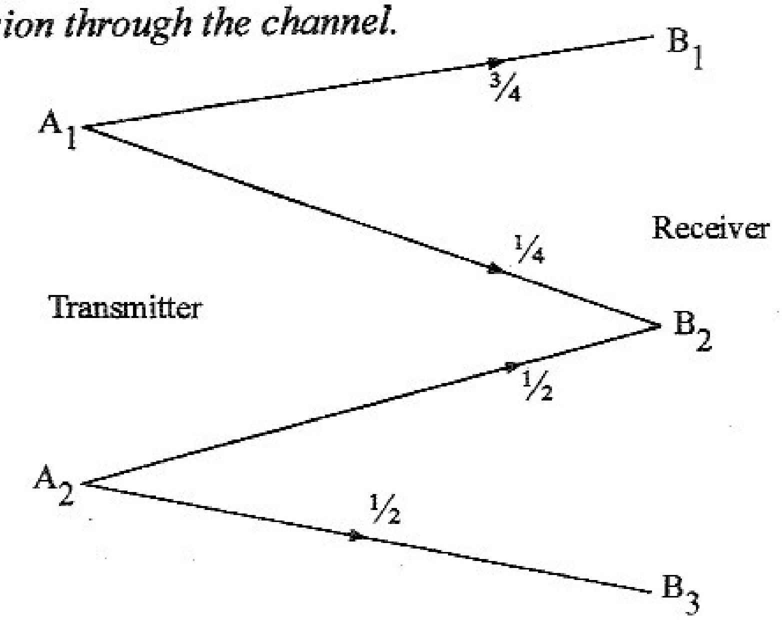


Fig. 1.19 Communication channel for problem 1.14

Solution

$$P(Y/X) = \begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

$$p(X, Y) = \begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} \frac{3}{8} & \frac{1}{8} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \end{array}$$

$$p(B_1) = \frac{3}{8}, \; p(B_2) = \frac{3}{8}, \; p(B_3) = \frac{1}{4}$$

$$H(Y) = -\frac{3}{8}\log\frac{3}{8} - \frac{3}{8}\log\frac{3}{8} - \frac{1}{4}\log\frac{1}{4}$$

$$= 1.5613 \text{ bits/symbol}$$

$$H(Y/X) = -\frac{3}{8}\log\frac{3}{4} - \frac{1}{8}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{2}$$

$$= 0.9056 \text{ bits/symbol}$$

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

$$= 1.5613 - 0.9056$$

$$I(X, Y) = 0.6557 \text{ bits/symbol}$$

Example 1.15

Given six messages m_1, m_2, m_3, m_4, m_5, m_6 with $p(m_1) = 1/3$, $p(m_2) = 1/4$, $p(m_3) = 1/8$, $p(m_4) = 1/8$, $p(m_5) = 1/12$, $p(m_6) = 1/12$. Find the Shannon - Fano code. Evaluate the coding Redundancy.

Solution

Message	Probability	codeword	length in symbols
m_1	$\frac{1}{3}$	00	2
	------- 2		
m_2	$\frac{1}{4}$	01	2
	------- 1		
m_3	$\frac{1}{8}$	100	3
	------- 4		
m_4	$\frac{1}{8}$	101	3
	------- 3		
m_5	$\frac{1}{12}$	110	3
	------- 5		
m_6	$\frac{1}{12}$	111	3

We first arrange the set in non-increasing probability order.

Average length of the code words is

$$\bar{L} = \left(\frac{1}{3}\right) \times 2 + \left(\frac{1}{4}\right) \times 2 + \left(\frac{1}{8}\right) \times 3 + \left(\frac{1}{8}\right) \times 3 + \left(\frac{1}{12}\right) \times 3 + \left(\frac{1}{12}\right) \times 3$$

$$= 2.41 \text{ symbols}$$

$$H(X) = -\left[\frac{1}{3}\log\frac{1}{3} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{8} + \frac{1}{8}\log\frac{1}{8} + \frac{1}{12}\log\frac{1}{12} + \frac{1}{12}\log\frac{1}{12}\right]$$

$$= 2.3758 \text{ bits/symbol}$$

$$\text{Coding efficiency } \eta_c = \frac{H(X)}{\bar{L}\log D} = \frac{2.3758}{2.41(\log_2 2)}$$

(D=2)

$$\eta_c = 0.985$$

The coding redundancy is $R_c = 1 - \eta_c = 0.015$

Eample 1.16

Given ten messages $m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8, m_9, m_{10}$ with corresponding probabilities $p(m_1) = 0.49$, $p(m_2) = 0.14$, $p(m_3) = 0.14$, $p(m_4) = 0.07$, $p(m_5) = 0.07$, $p(m_6) = 0.04$, $p(m_7) = 0.02$, $p(m_8) = 0.02$, $p(m_9) = 0.005$, $p(m_{10}) = 0.005$. Find the Shannon-Fano Code and calculate the coding Redundancy.

Solution

We first arrange the set in non-increasing probability order.

Messages	Probability	Codewords	Lengths
m_1	0.49	0	1
	------- 1		
m_2	0.14	100	3
	------- 3		
m_3	0.14	101	3
	------- 2		
m_4	0.07	1100	4
	------- 5		
m_5	0.07	1101	4
	------- 4		
m_6	0.04	1110	4
	------- 6		
m_7	0.02	11110	5
	------ -7		
m_8	0.02	111110	6
	------- 8		
m_9	0.005	1111110	7
	----- -9		
m_{10}	0.005	1111111	7

$$H(X) = -0.49 \log 0.49 - 0.14 \log 0.14 - 0.14 \log 0.14 - 0.07 \log 0.07 - 0.07 \log 0.07 - 0.04 \log 0.04 - 0.02 \log 0.02 - 0.02 \log 0.02 - 0.005 \log 0.005 - 0.005 \log 0.005$$

$$= 2.3236 \text{ bits/symbol}$$

$$\overline{L} = 0.49 \times 1 + 0.14 \times 3 + 0.14 \times 3 + 0.07 \times 4 + 0.07 \times 4 + 0.04 \times 4 + 0.02 \times 5 + 0.02 \times 6 + 0.005 \times 7 + 0.005 \times 7$$

$$= 2.34 \text{ symbols.}$$

Hence
$$\eta_c = \frac{H(X)}{\bar{L}\log D} = \frac{2.3236}{2.34} = 0.993$$

$$R_c = 1 - \eta_c = 0.007.$$

Example 1.17

Apply the Shannon - Fano encoding procedure to the follwing message ensemble :

$$[X] = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]$$
$$[P] = [0.49, 0.14, 0.14, 0.07, 0.07, 0.04, 0.02, 0.02, 0.01]$$

Message	Probability	Codeword	Length
x_1	0.49	0	1
	-------- 1		
x_2	0.14	100	3
	-------- 3		
x_3	0.14	101	3
	-------- 2		
x_4	0.07	1100	4
	-------- 5		
x_5	0.07	1101	4
	-------- 4		
x_6	0.04	1110	4
	-------- 6		
x_7	0.02	11110	5
	-------- 7		
x_8	0.02	111110	6
	-------- 8		
x_9	0.01	111111	6

$$H(X) = -\ [0.49 \log 0.49 + 0.28 \log 0.14 + 0.14 \log 0.07 + 0.04 \log 0.04 + 0.04 \log 0.02 + 0.01 \log 0.01]$$

$$\bar{L} = 0.49 \times 1 + 0.28 \times 3 + 0.18 \times 4 + 0.02 \times 5 + 0.03 \times 6 = 2.33$$

$$\text{Efficiency} = \frac{H(X)}{2.33} = 0.993$$

Example 1.18

Apply a ternary partitioning technique (similar to the Shannon-Fano procedure) for encoding the following messages in codes using an alphabet [0, 1, 2]. Find $\bar{L}$ and the code efficiency.

$$\{m\} = \{m_1, m_2, m_3, m_4, m_5, m_6\}$$

$$p\{m\} = \left\{\frac{3}{8}, \frac{1}{6}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{12}\right\}$$

Solution

Message	Probability	Codeword	Length
m_1	0.375	0	1
	------- 1		
m_2	0.167	10	2
	------- 2		
m_3	0.125	11	2
	------- 1		
m_4	0.125	20	2
	------- 3		
m_5	0.125	21	2
	------- 3		
m_6	0.083	22	2

The probabaility of the occurrence of 0, 1 and 2 can be directly computed

$$p\{0\} = \frac{\sum_{k=1}^{6} p\{m_k\} C_{ko}}{\sum_{k=1}^{6} p\{m_k\}.n_k}$$

$$= \frac{[(0.375).1 + (0.167).1 + (0.125).0 + (0.125).1 + (0.125)0 + (0.083).0]}{[(0.375).1 + (0.167).2 + (0.125).2 + (0.125).2 + (0.125).2 + (0.083).2]}$$

$$p\{0\} = \frac{16}{39}$$

$$p\{1\} = \frac{13}{39}$$

$$p\{2\} = \frac{10}{39}$$

$$\bar{L} = \sum p(m_i).n_i = \frac{39}{24} = 1.62$$

$$H(X) = 2.388$$

$$\text{Efficiency} = \frac{2.388}{1.62 \times 1.584} = 0.94$$

Example 1.19

Find the smallest number of letters in the alphabet (number D) for devising a code with a prefix property such that

$$[W] = [0, 3, 0, 5]$$

Devise such a code.

Solution

If n_i = word length
and W_i = number of encoded messages of length n_i,
the condition for the existence of desired code is

$$\sum_{j=1}^{m} W_j D^{-j} \leq 1$$

$$3D^{-2} + 5D^{-4} \leq 1$$

The inequality is satisfied for

$$D^2 \geq \frac{3 + \sqrt{29}}{2}$$

The smallest permissible value is D = 3. That is, no binary code can be devised with the above constraint. To devise such a code, let the alphabet be [0, 1, 2] ; then one of the several encoding procedures is as follows.

m_1	00
m_2	01
m_3	02
m_4	1000
m_5	1001
m_6	1002
m_7	2000
m_8	2222

Example 1.20

Determine whether or not each of the following codes is uniquely decipherble. If a code is not uniquely decipherable, construct an ambiguous sequence.

Solution

(a)

	x_1	010
	x_2	0001
	x_3	0110
	x_4	1100
	x_5	00011
	x_6	00110
	x_7	11110
	x_8	101011

(b)

	x_1	abc
	x_2	abcd
	x_3	e
	x_4	dba
	x_5	bace
	x_6	ceac
	x_7	ceab
	x_8	eabd

Solution

Referring to Appendix-I, we form the following table.

(a)

S_0	S_1	S_2	S_3	S_4	S_5	S_6
010	1	100	11	00	01	0
0001		1110		110	011	10
0110		01011			110	001
1100					0	110
00011						0011
00110						code word
11110						
101011						

The code is not uniquely decipherable as S_6 contains 0110 which is a code word in S_0.

$A_0 = 0001 \;\; A_1 = 1 \;\; A_2 = 01011 \;\; A_3 = 11 \;\; A_4 = 110 \;\; A_5 = 0 \;\; A_6 = 0110$

$W_0 = 00011 \;\; W_1 = 101011 \;\; W_2 = 010 \;\; W_3 = 11110 \;\; W_4 = 1100 \;\; W_5 = 00110$
$W_6 = 0110$

Ambiguous Sequence $= A_0 W_1 W_4 W_6 = 0001 \;\; 101011 \;\; 1100 \;\; 0110$

$= W_0 W_2 W_3 W_5 = 00011 \;\; 010 \;\; 11110 \;\; 00110$

(b)

S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
abc	d	ba	ce	ac	cd	eac	ac	c	eac	ac
abcd	abd			ab	c	eab	ab	cd	eab	ab
e							d	ba	ce	d
dba										
bace										
ceac										
ceab										
eabd										

$S_7 = S_{10}\,; \; S_i = S_{i+3}, \;\; i \geq 7$

None of the sets S_i contains a code word, so the code is uniquely decipherable.

Example 1.21

The output of a discrete source,

$[X] = [x_1, x_2, x_3, x_4, x_5, x_6]$

$P[X] = [2^{-1}, 2^{-2}, 2^{-4}, 2^{-4}, 2^{-4}, 2^{-4}]$

is encoded in the following six ways:

	C_1	C_2	C_3 ✓	C_4 ✓	C_5 ✓	C_6
x_1	0	1	0	111	1	0
x_2	10	011	10	110	01	01
x_3	110	010	110	101	0011	011
x_4	1110	001	1110	100	0010	0111
x_5	1011	000	11110	011	0001	01111
x_6	1101	110	111110	010	0000	011111

(a) Determine which of those codes are uniquely decipherable.

(b) Determine those which have the prefix property.

(c) Find the average length of each uniquely decipherable code.

(d) Does any one of the above codes give minimum average length?

Solution

(a)

$C_1(S_0)$	S_1	S_2		S_3	S_4
0	11	$\boxed{0}$	code word	11	0
10	1	10		10	
110		01		01	
1110					
1011					
1101					

$S_2 = S_4$; $S_i = S_{i+2}$, $i \geq 2$

Therfore, S_2 contains a code word of S_0.

Hence the codes C_1 are not uniquely decipherable.

$C_2(S_0)$	S_1	S_2		S_3	S_4	S_5	S_6
1	10	$\boxed{0}$	code word	11	0	11	0
011		1		10		10	
010				01		01	
001				00		00	
000				10			
110							

$S_4 = S_6$; $S_i = S_{i+2}$, $i \geq 4$

Therefore, S_2 contains a code word of S_0.

Hence the codes C_2 are not uniquely decipherable.

Similarly one finds that C_3, C_4, C_5 and C_6 are uniquely decipherable.

While it is clear that C_1 and C_2 are not uniquely decipherable, we wish to apply McMillan's realizability criterion

For C_1: $\quad 2^{-1} + 2^{-2} + 2^{-3} + 3.\, 2^{-4} > 1$

For C_2: $\quad 2^{-1} + 5.2^{-3} > 1$

Thus such uniquely decipherable codes cannot exist.

(b) C_3, C_4, C_5 have prefix property but C_6 does not have such property

(c) $\bar{L}_3 = \frac{1}{2} + \frac{2}{4} + \frac{3}{16} + \frac{4}{16} + \frac{5}{16} + \frac{6}{16} = 2\frac{1}{8}$

$\bar{L}_4 = 3 \quad \bar{L}_5 = 2 \quad \bar{L}_6 = 2\frac{1}{8}$

(d) Entropy of the source is

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{4}{16}\log\frac{1}{16} = 2 \text{ bits/symbol}$$

We know $$\bar{L} \geq \frac{H(X)}{\log D}$$

According to this, the average length of a uniquely decipherable code cannot be less than H(X).

Thus C_5 is a code that achieves minimum average length.

Example 1.22

Apply Shannon's encoding procedure to the following message ensemble:

$[X] = [x_1, x_2, x_3, x_4]$

$[P] = [0.4, 0.3, 0.2, 0.1]$

Solution:

Step 1 : $0.4 > 0.3 > 0.2 > 0.1$

Step 2 : $\alpha_1 = 0 \quad \alpha_2 = 0.4 \quad \alpha_3 = 0.4 + 0.3 = 0.7 \quad \alpha_4 = 0.7 + 0.2 = 0.9$

Step 3 : $0.4 \geq 2^{-2},\ n_1 = 2$

$0.3 \geq 2^{-2},\ n_2 = 2$

$0.2 \geq 2^{-3},\ n_3 = 3$

$0.1 \geq 2^{-4},\ n_4 = 4$

Step 4 : $\alpha_1 = 00$

$\alpha_2 = 0.4 = 01/1$

$\alpha_3 = 0.7 = 101/1$

$\alpha_4 = 0.9 = 1110/$

x_1	00
x_2	01
x_3	101
x_4	1110

Example 1.23

Given the following set of messages and their corresponding transmission probabilities

$$[m_1, m_2, m_3]$$

$$\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$$

(a) Construct a binary code satisfying the prefix condition and having the minimum possible average length of encoded digits. Compute the efficiency of the code.

(b) For a source transmitting messages of second-order extension,

$$\begin{bmatrix} m_1m_1 & m_1m_2 & m_1m_3 \\ m_2m_1 & m_2m_2 & m_2m_3 \\ m_3m_1 & m_3m_2 & m_3m_3 \end{bmatrix}$$

Construct a binary code with prefix property and minimum average length and compute its efficiency.

Solution

(a) If the binary code must have the prefix property, then we assign the following code :

m_1	0
m_2	10
m_3	11

The average length of the code word is

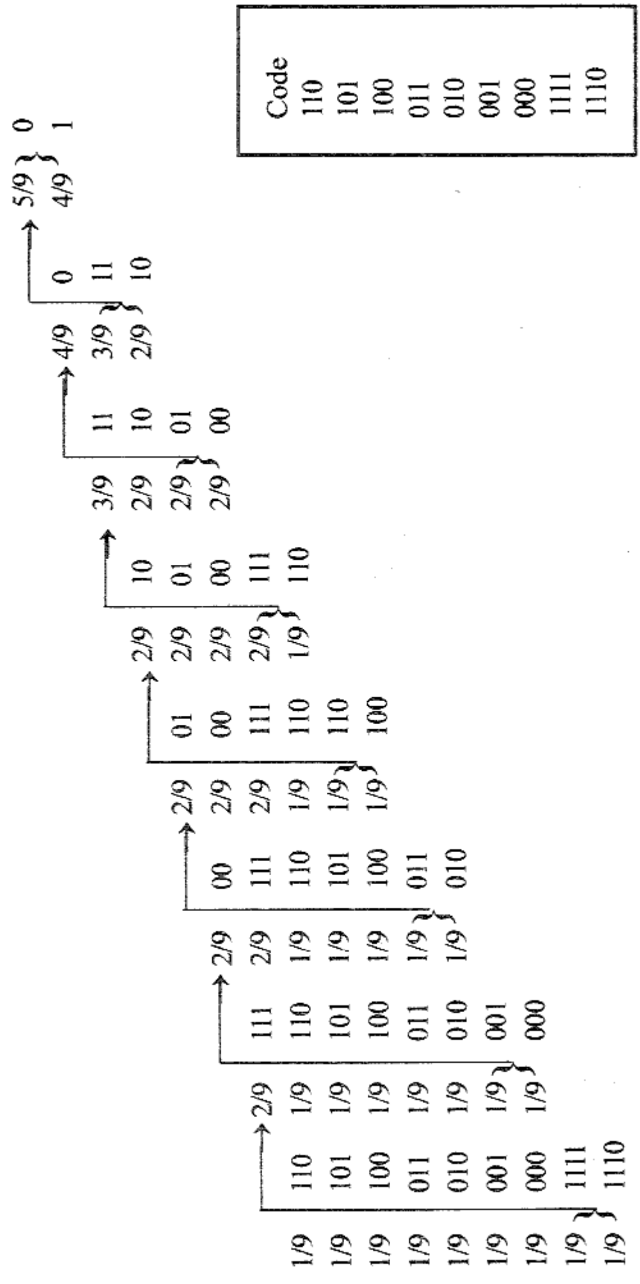
$$\frac{1}{3}(1)+\frac{1}{3}(2)+\frac{1}{3}(2)=\frac{5}{3}\text{ binits.}$$

0 and 1 appear with probabilities of $\frac{2}{5}$ and $\frac{3}{5}$ respectively.

$$\text{Efficiency} = \frac{\log 3}{\frac{5}{3}\log 2} = 0.95$$

Shannon's encoding also leads to the same result.

(b) We construct a Huffman code which is shown in the following page.


Code
110
101
100
011
010
001
000
1111
1110

Example 1.24

Apply Huffmann's encoding procedure to the following message ensemble and determine the average length of the encoded message.

$$\{X\} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$p\{X\} = \{0.18, 0.17, 0.16, 0.15, 0.10, 0.08, 0.05, 0.05, 0.04, 0.02\}$$

The encoding alphabet is $\{D\} = \{0, 1, 2, 3\}$

Solution:

					→0.49	0
x_1	0.18	1	0.18	1	0.18	1
x_2	0.17	2	0.17	2	0.17	2
			→ 0.16	3	0.16	3
x_3	0.16	00	0.16	00		
x_4	0.15	01	0.15	01		
x_5	0.10	02	0.10	02		
x_6	0.08	03	0.08	03		
x_7	0.05	30				
x_8	0.05	31				
x_9	0.04	32				
x_{10}	0.02	33				

	Code
x_1	1
x_2	2
x_3	00
x_4	01
x_4	02
x_5	02
x_6	03
x_7	30
x_8	31
x_9	32
x_{10}	33

$$\overline{L} = (0.18 \times 1) + (0.17 \times 1) + (0.16 \times 2) + (0.15 \times 2) + (0.18 \times 2) + (0.05 \times 2) + (0.05 \times 2) + (0.04 \times 2) + (0.02 \times 2)$$

$$\overline{L} = 1.65.$$

Example 1.25

a) For a binary code below, let N(k) be the number of messages that can be formed using exactly K code characters. For example, N(1) =1. (that is x_1*),* $N(2) = (x_1 x_2, x_2, x_3)$*,* $N(3) = 5\ (x_1 x_1 x_1, x_1 x_2, x_1 x_3, x_2 x_1, x_3 x_1)$*. Find a general expression for N(k) (k = 1, 2, . . .)*

x_1	*0*
x_2	*10*
x_3	*11*

(b) Repeat part (a) for the code below

x_1	*0*
x_2	*10*
x_3	*110*
x_4	*111*

Solution

Let w_i be the number of code words of length i ($w_1 = 1, w_2 = 2, w_n = 0, n \geq 3$). Then a message whose coded form has exactly k letters must begin with a code word of length 1 or a code word of length 2; hence

$$N(k) = w_1 N(k-1) + w_2 N(k-2), K \geq 3$$

Now $N(2) = w_1 N(1) + w_2$; hence if we define $N(0) = 1$, the above equation is valid for $k \geq 2$.

Thus we must solve the linear homogeneous difference equation

$N(k+2) - w_1 \ N(k+2) - w_2 N(k) = 0, k = 0, 1, \ldots.$

subject to $N(0) = 1, N(1) = 1$.

Assuming a solution of the form $N(k) = \lambda^K$, we obtain

$\lambda^{K+1} - w_1 \ \lambda^{K+1} - w_2 \lambda^K = 0$ or $\lambda^K[\lambda^2 - \lambda - 2] = 0.1$

The two nonzero roots are $\lambda_1 = 2, \lambda_2 = -1$.

Hence, $N(k) = A. 2^K + B(-1)^K, K = 0, 1, \ldots$

Since $N(0) = A + B = 1, N(1) = 2A - B = 1$, we have

$$A = \frac{2}{3}, \ B = \frac{1}{3}.$$

$$N(K) = \frac{2}{3} 2^K + \frac{1}{3}(-1)^K, k = 0, 1 \ldots$$

(b) As in (a), $N(k) = w_1 N(k-1) + w_2 N(k-2) + w_3 N(k-3)$.

$$(w_1 = 1, w_2 = 2, w_3 = 2, w_n = 0, n \geq 4)$$

or $N(k+3) - N(k+2) - N(k+1) - 2N(k) = 0, k \geq 0$

The assumption $N(K) = \lambda^k$ yields the characteristic equation

$$\lambda^3 - \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda^2 + \lambda + 1) = 0$$

with roots

$$\lambda_1 = 2, \ \lambda_2 = -\frac{1}{2} - \frac{1}{2} i\sqrt{3} = e^{i\left(\frac{2\pi}{3}\right)}$$