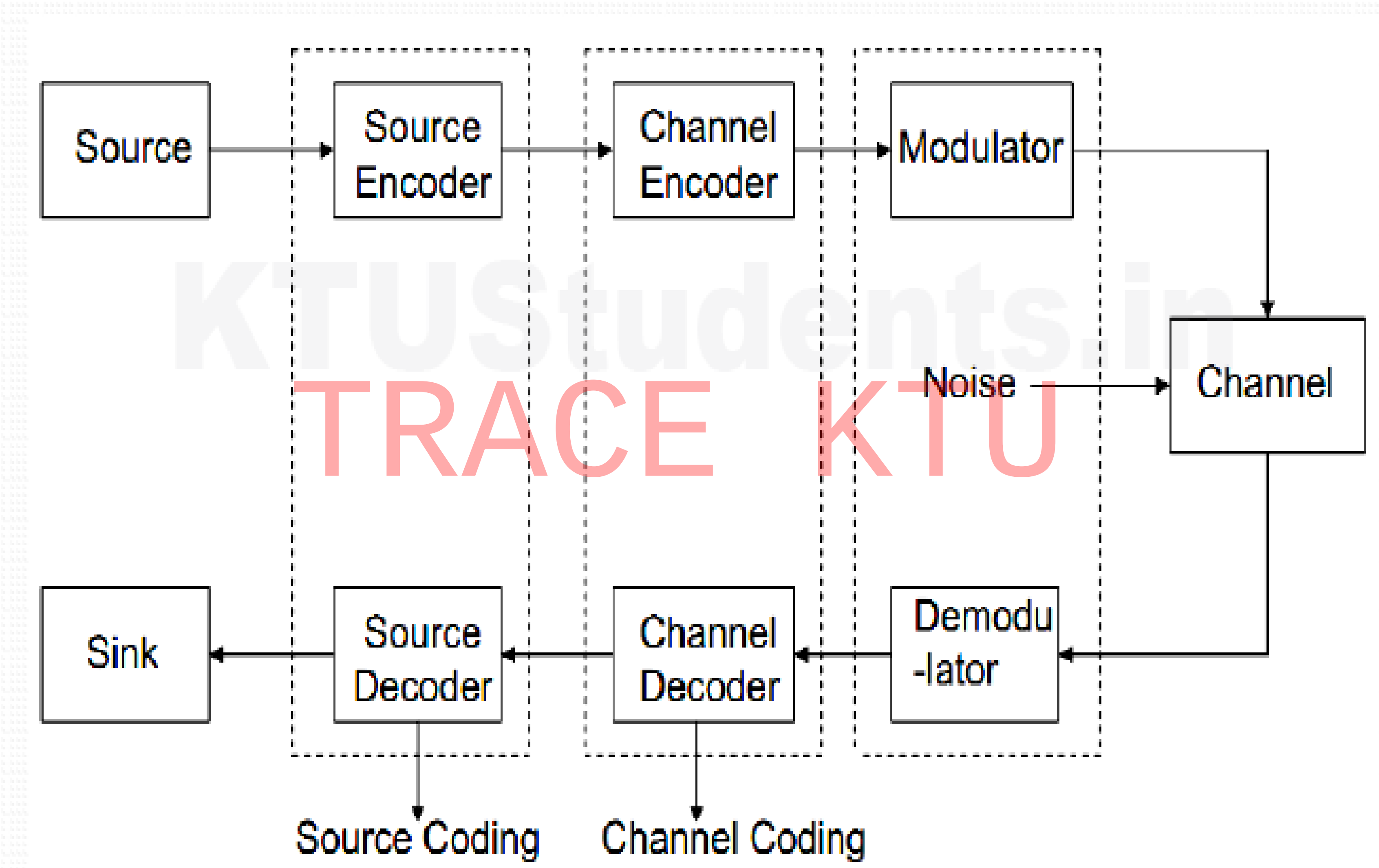
ECT306 INFORMATION THEORY & CODING

MODULE 1
Entropy, Sources and Source Coding

- Entropy, Properties of Entropy, Joint and Conditional Entropy, Mutual Information, Properties of Mutual Information.
- Discrete memoryless sources, Source code, Average length of source code, Bounds on average length, Uniquely decodable and prefix-free source codes. Kraft Inequality (with proof), Huffman code. Shannon's source coding theorem (both achievability and converse) and operational meaning of entropy.

Digital Communication System



Digital Communication System

Source encoding: converts source output to bits. Source input can be voice, video, text, sensor output.

Channel coding: adds extra bits (redundancy) to data which is to be transmitted over the channel.

- This redundancy helps reduce the errors which can be introduced in transmitted bits due to channel noise.

Information Theory

Information theory answers two fundamental questions in communications

- What is the ultimate data compression?
- What is the ultimate transmission rate?

Concept of information

- A flip of a coin with two heads? Does it convey any information?
- A source produces successive bits of pi:3,1,4,1,5,9,2,6. Does it convey any information?
- There is no uncertainty in the source.
- Shannon’s Information Theory regards only those symbols as information that are not predictable.

Concept of information

- Box A

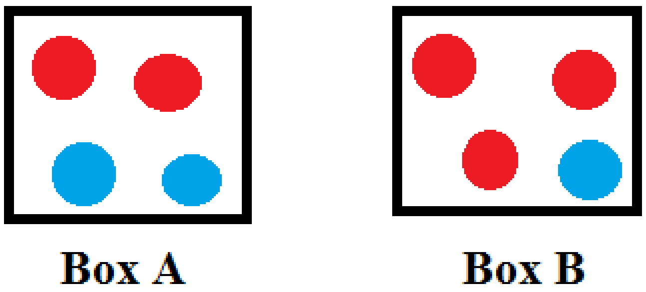
P(Getting red ball)=1/2

P(Getting blue ball)=1/2

- Box B

P(Getting red ball)=3/4

P(Getting blue ball)=1/4



- The information conveyed by an experiment is depends on probability of occurrence.

Concept of information

- The source output is modeled as a discrete random variable, S, which takes on symbols from a fixed finite alphabet

$$S = \{s_0, s_1, s_2, \dots \dots, s_{K-1}\}$$

with probabilities

$$P(S = s_k) = p_k, \qquad k = 0,1, \dots \dots, K-1$$

- This set of probabilities must satisfy the condition

$$\sum_{k=0}^{K-1} p_k = 1$$

Concept of information

- The amount of information is related to the inverse of the probability of occurrence.
- We define the amount of information gained after observing the event $S = s_k$, which occurs with probability p_k , as the logarithmic function

$$I(s_k) = log_2\left(\frac{1}{p_k}\right)$$

- It is also known as *self information.*

Properties of Self Information

1. $I(s_k) = 0 \quad \text{for} \quad p_k = 1$

   If we are absolutely certain of the outcome of an event, even before it occurs, there is no information gained.

2. $I(s_k) \geq 0 \quad \text{for} \quad 0 \leq p_k \leq 1$

   The occurrence of an event $S = s_k$ either provides some or no information, but never brings about a loss of information.

3. $I(s_k) > I(s_i) \quad \text{for} \quad p_k < p_i$

   That is, the less probable an event is, the more information we gain when it occurs.

4. $I(s_k s_i) = I(s_k) + I(s_i)$ if s_k and s_i are statistically independent.

Units of Information

- If the base is 3, unit of information is **Triples.**
- If the base is 4, unit of information is **Quadruples.**
- If the base is 10, unit will be **Hartley's.**
- If the base is e (natural logarithm), unit is **Nats.**
- In the context of digital communication, the base is taken as 2.
- The resulting unit of information is called the **bit.**
- One bit is the amount of information that we gain when one of two possible and equally likely events occurs.

When $p_k = 1/2$, we have $I(s_k) = 1\ bit$

Units of Information

1 Hartley = 3.322 bits.

1 Nat = 1.443 bits.

1 Triple = 1.585 bits.

Entropy or Average Information

- $I(s_k)$ is a discrete random variable. Mean of $I(s_k)$ over the source alphabet S is called entropy of the discrete memoryless source.

$$H(S) = E[I(s_k)]$$
$$= \sum_{k=0}^{K-1} p_k I(s_k)$$
$$= \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

- Entropy is a measure of the average information content per source symbol or it is a measure of uncertainty (randomness) about S. It is average self information.

Question

- A source S emits symbols S_o, S_1, S_2 and S_3 with probabilities ½, 1/4, 1/8, 1/8 respectively. Find entropy of S.

$$H(S) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

$$H(S) = -\sum_{k=0}^{K-1} p_k \log_2(p_k)$$

$$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8}$$

$$= \frac{7}{4}\text{ bits}$$

Previous Univ Question (3 Mark)

- A source emits one of four symbols S_o, S_1, S_2 and S_3 with probabilities 1/3, 1/6, 1/4, 1/4 respectively. The successive symbols emitted by the source are statistically independent. Calculate the entropy of the source.

Properties of Entropy

Consider a discrete memoryless source. The entropy $H(S)$of such a source is bounded as follows:

$$0 \le H(S) \le log_2 K$$

where K is the *radix* (number of symbols) of the alphabet $S$of the source. Furthermore we may make two statements:

1. $H(S) = 0$, if and only if the probability $p_k = 1$ for some k and the remaining probabilities in the set are all zero; this lower bound on entropy corresponds to *no uncertainty*.
2. $H(S) = log_2 K$, if and only if $p_k = 1/K$ for all k (i.e., all the symbols in the alphabet S are *equiprobable);* this upper bound on entropy corresponds to *maximum uncertainty*.

The entropy of a source is maximum when the symbols are equiprobable.

Prrof:

When entropy H is maximum,

$$\frac{dH}{dp} = 0$$

Consider two symbols with probabilities p and 1-p.

The entropy $H = plog\left(\frac{1}{p}\right) + (1-p)\log\left(\frac{1}{1-p}\right)$

$$= -plogp - (1-p)\log(1-p)$$

$$\frac{dH}{dp} = -p\left(\frac{1}{p}\right) + \log p \times -1 - \left[(1-p)\left(\frac{1}{1-p}\right) \times -1 - \log(1-p) \times -1\right]$$

$$= -1 - \log(p) - [-1 + \log(1-p)]$$

$$= -\log(p) + \log(1-p)$$

Equating to zero, i.e,

$$\frac{dH}{dp} = 0$$

$$0 = -\log(p) + \log(1-p)$$

$$\log(p) = \log(1-p)$$

Take antilog on both sides

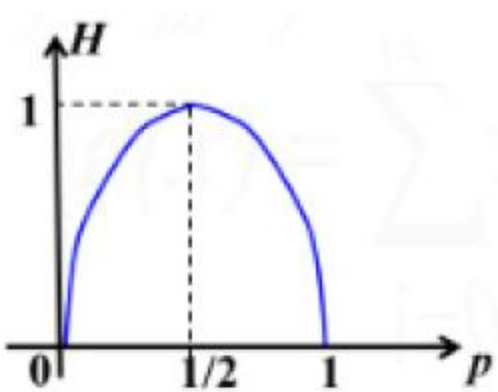
$$p = 1 - p$$

$$2p = 1$$

$$p = 1/2$$

i.e, entropy is maximum (value 1) when probability is ½.

i.e., the maximum uncertainity occurs when symbols are equiprobable.



When all the probabilities are equal,

$$H = log_2 K$$

Proof:

Consider a source emits K equiprobable symbols with probability 1/K. Then

$$H = \sum_{k=0}^{K-1} p_k \, log_2\left(\frac{1}{p_k}\right)$$

$$H = \sum_{k=0}^{K-1} \frac{1}{K} log_2 K$$

$$= K \frac{1}{K} log_2 K$$

$$= log_2 K$$

Rate of Information

- Suppose the symbols emitted by the source by a fixed rate *r symbols/sec*.
- The average source information rate R in bits/sec is defined as the product of average information content per symbol H and the message symbol rate r .
- Rate of information $R = rH$ *bits / sec*

  where r is the symbol rate

  H is the average information content per symbol or Entropy.

Questions

Example 1:

Consider a binary source, s1 and s2 with probabilities 1/256 and 255/256. Find the entropy.

Ans:0.036 bits/symbol

Example 2:

An event has 6 possible outcomes with probabilities p1=1/2, p2=1/4, p3=1/8, p4=1/16, p5=1/32, p6=1/32. Find the entropy of the system. Also find the rate of information if there are 16 outcomes/sec.

Questions

Example 3: The output of an information source consist of 150 symbols, 32 of which occur with a probability of 1/64 and the remaining 118 occurs with a probability of 1/236. The source emits 2000 symbols/sec. Find the information rate of source.

Example 4: An analog signal is bandlimited to 500 Hz and is sampled at Nyquist rate. The samples are quantized into 4 levels. The quantization levels are assumed to be independent and occur with probabilities p1=p4=1/8, p2=p3=3/8. Find the information rate of the source.

Entropy of Extended Source

- Consider blocks rather than individual symbols.
- Each block consist of n successive source symbols.
- Each block is produced by an extended source with a source alphabet S^n that has K^n distinct blocks, where K is the number of distinct symbols.
- The Entropy of extended source $H(S^n)$ is equal to n times $H(S)$, the entropy of original source.

$$H(S^n) = nH(S)$$

- $H(S^n)$ is also known as n^{th} order entropy.

Example

Consider a discrete memoryless source with source alphabet S = $\{s_0, s_1, s_2\}$ with respective probabilities

$P_0=1/4$

$P_1=1/4$

$P_2=1/2$

Find first order and second order entropies.

Ans:

First order Entropy

$$H(\mathscr{S}) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right)$$

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

$$= \frac{3}{2} \text{ bits}$$

Second order extension of sources

Symbols of $\mathcal{S}^2$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Corresponding sequences of symbols of $\mathcal{S}$	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Probability $p(\sigma_i)$, $i = 0, 1, \ldots, 8$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

Entropy of second order extension or second order entropy is

$$H(\mathcal{S}^2) = \sum_{i=0}^{8} p(\sigma_i) \log_2 \frac{1}{p(\sigma_i)}$$

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

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

$$= 3 \text{ bits}$$

Previous Univ Question (8 Mark)

- Consider a source with alphabet, S={x1, x2}, with respective probabilities 1/4 and 3/4. Determine the entropy, H(S) of the source. Write the symbols of the second order extension of S, i.e., S^2 and determine its entropy, $H(S^2)$. Verify that $H(S^2) = 2 H(S)$.

Entropy – 0.8113 bits/symbol

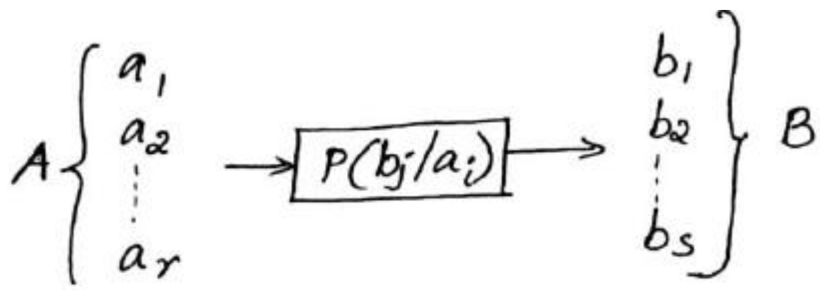
Extended symbols	Probabilities
x_1x_1	1/16
x_1x_2	3/16
x_2x_1	3/16
x_2x_2	9/16

Entropy – 1.6225bits/symbol

Communication Channels

- Medium through which the coded signals generated by an information source are transmitted.
- The input to the channel is a symbol belongs to an alphabet '*A*' with '*r*' symbols.
- The output of the channel is a symbol belongs to some other alphabet '*B*' with '*s*' symbols.
- Due to error in the channel, the output symbol may be different from the input symbol during any symbol interval.

Representation of a Channel



- These conditional probabilities come into existence due to the presence of noise in the channel.
- Because of this noise, there will be some amount of uncertainty about the reception of any symbol.
- So different number of symbols 's' at the receiver from 'r' symbols at the transmitter.

CHANNEL MATRIX

- $r \times s$ number of conditional probabilities which are represented in a "matrix" form with all the input symbols represented row-wise and output symbols column-wise.

$$P(B/A) = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array} \begin{array}{c} \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} \\ \begin{bmatrix} P(b_1/a_1) & P(b_2/a_1) & \cdots & P(b_s/a_1) \\ P(b_1/a_2) & P(b_2/a_2) & \cdots & P(b_s/a_2) \\ \vdots & & & \\ P(b_1/a_r) & P(b_2/a_r) & \cdots & P(b_s/a_r) \end{bmatrix} \end{array}$$

- The sum of all the elements in any row of the channel matrix is equal to unity.

- Knowing the channel matrix elements and the input probabilities, the probabilities of the output symbol can be found using the “theorem of total probability”.

$$P(b_1) = P\left({}^{b_1}/_{a_1}\right)P(a_1) + P\left({}^{b_1}/_{a_2}\right)P(a_2) + \cdots \ldots\ldots +P\left({}^{b_1}/_{a_r}\right)P(a_r)$$

$$P(b_2) = P\left({}^{b_2}/_{a_1}\right)P(a_1) + P\left({}^{b_2}/_{a_2}\right)P(a_2) + \cdots \ldots\ldots +P\left({}^{b_2}/_{a_r}\right)P(a_r)$$

.

.

$$P(b_s) = P\left({}^{b_s}/_{a_1}\right)P(a_1) + P\left({}^{b_s}/_{a_2}\right)P(a_2) + \cdots \ldots\ldots +P\left({}^{b_s}/_{a_r}\right)P(a_r)$$

JOINT PROBABILITY MATRIX (JPM)

- The input conditional probability can be found by using Baye's Rule.

$$P\left({a_i}/{b_j}\right) = \frac{P\left({b_j}/{a_i}\right)P(a_i)}{P(b_j)}$$

- The joint probability between any input symbol a_i and any output symbol b_j is given by,

$$P(a_i, b_j) = P\left({b_j}/{a_i}\right)P(a_i) = P\left({a_i}/{b_j}\right)P(b_j)$$

JOINT PROBABILITY MATRIX (JPM)

- The matrix whose elements are the various joint probabilities between input and output symbols is called JOINT PROBABILITY MATRIX (JPM)

$$P(a_i, b_j) = P(A, B) = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array} \begin{array}{c} \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} \\ \begin{bmatrix} P(a_1, b_1) & P(a_1, b_2) & \cdots & P(a_1, b_s) \\ P(a_2, b_1) & P(a_2, b_2) & \cdots & P(a_2, b_s) \\ \vdots & \vdots & & \vdots \\ P(a_r, b_1) & P(a_r, b_2) & \cdots & P(a_r, b_s) \end{bmatrix} \end{array}$$

Properties of JPM

- By adding the elements of JPM column wise, we can obtain the probability of output symbols.
- By adding the elements of JPM row wise, we can obtain the probability of input symbols.
- The sum of all the elements of JPM is equal to unity.

Joint Entropy

- It is the average information per pairs of transmitted and received symbols (x_i, y_j) or the average uncertainty of the communication link as a whole.

$$H(X,Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\, log_2 P(x_i, y_j)$$

Conditional Entropy

$$H\left(Y/_{X}\right) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\, log_2 P\left({}^{y_j}/_{x_i}\right)$$

- It is the entropy of the received symbol when transmitted state is known, or it is the average uncertainty of channel output given that X was transmitted.

$$H\left(X/_{Y}\right) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\, log_2 P\left({}^{x_i}/_{y_j}\right)$$

- It is the entropy of the source when the state of the receiver is known.

Relationships among various entropies

$$H(X,Y) = H\left(X/_{Y}\right) + H(Y)$$

$$H(X,Y) = H\left(Y/_{X}\right) + H(X)$$

Mutual Information

- It is defined as the amount of information transferred where $\boldsymbol{x_i}$ is transmitted and $\boldsymbol{y_j}$ is received.

$$I(x_i, y_j) = log_2 \left[\frac{P\left({x_i}/{y_j} \right)}{P(x_i)} \right]$$

Average Mutual Information

- It is represented by $I(X;Y)$
- It is calculated in *bits/symbol*
- It is defined as amount of source information gained per received symbol.

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, I(x_i, y_j)$$

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left[\frac{P\left({x_i}/{y_j}\right)}{P(x_i)}\right]$$

We have the relation

$$P\left({x_i}/{y_j}\right) = \frac{P\left(x_i, y_j\right)}{P\left(y_j\right)}$$

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left(x_i, y_j\right)}{P(x_i)P(y_j)}\right)$$

Similarly

$$I(Y;X) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left({y_j}/{x_i}\right)}{P\left(y_j\right)}\right)$$

Properties of Mutual Information

1. Mutual Information is symmetry.

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

2. Mutual information is always non-negative

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

3. Mutual information can be expressed as Entropies

$$I(X;Y) = H(X) - H\left({}^{X}/_{Y}\right)$$

$$I(X;Y) = H(Y) - H\left({}^{Y}/_{X}\right)$$

4. Mutual information is related to Joint Entropy as

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

Proof of Property 1 $I(X;Y) = I(Y;X)$

$$P(x_i, y_j) = P\left({x_i}/{y_j}\right) P(y_j) \qquad (1)$$

$$P(x_i, y_j) = P\left({y_j}/{x_i}\right) P(x_i) \qquad (2)$$

$$P\left({x_i}/{y_j}\right) P(y_j) = P\left({y_j}/{x_i}\right) P(x_i)$$

$$\frac{P\left({x_i}/{y_j}\right)}{P(x_i)} = \frac{P\left({y_j}/{x_i}\right)}{P(y_j)} \qquad (3)$$

$$I(X;Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left({x_i}/{y_j}\right)}{P(x_i)}\right) \qquad (4)$$

$$I(Y;X) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left(y_j/x_i\right)}{P(y_j)}\right) \qquad (5)$$

Substitute eqn (3) in (5)

$$I(Y;X) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left(x_i/y_j\right)}{P(x_i)}\right) \qquad (6)$$

From (4) and (6) we get

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

Hence proved.

Proof of Property 2 $I(X;Y) \geq 0$

We know

$$I(X;Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) \, log_2 \left(\frac{P(x_i, y_j)}{P(x_i) P(y_j)} \right)$$

Take minus sign outside

$$I(X;Y) = - \sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) \, log_2 \left(\frac{P(x_i) P(y_j)}{P(x_i, y_j)} \right)$$

$$-I(X;Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) \, log_2 \left(\frac{P(x_i) P(y_j)}{P(x_i, y_j)} \right)$$

We know the fundamental inequality

$$\sum_{k=0}^{K-1} P_k \, log_2\left(\frac{q_k}{P_k}\right) \leq 0$$

Using this inequality in the previous equation we can write

$$-I(X;Y) \leq 0$$

or

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

Hence Proved

Proof of Property 3

$$I(X;Y) = H(X) - H\left(X/_Y\right)$$

$$I(X;Y) = H(Y) - H\left(Y/_X\right)$$

$$H\left(X/_Y\right) = \sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\, log_2\left(\frac{1}{P\left(x_i/_{y_j}\right)}\right)$$

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{P\left(x_i/_{y_j}\right)}{P(x_i)}\right)$$

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{1}{P(x_i)}\right) - \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) log_2\left(\frac{1}{P\left(x_i/_{y_j}\right)}\right)$$

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j)\, log_2\left(\frac{1}{P(x_i)}\right) - H\left(X/_Y\right)$$

By the property of JPM

$$\sum_{j=1}^{n} P\left(x_i, y_j\right) = P(x_i)$$

$$I(X;Y) = \sum_{i=1}^{m} P(x_i)\, log_2\left(\frac{1}{P(x_i)}\right) - H\left({}^{X}/_{Y}\right)$$

$$I(X;Y) = H(X) - H\left({}^{X}/_{Y}\right)$$

Hence proved

Proof of Property 4 $I(X;Y) = H(X) + H(Y) - H(X,Y)$

$$H(X,Y) = H\left(X/_Y\right) + H(Y) \qquad (1)$$

$$I(X;Y) = H(X) - H\left(X/_Y\right) \qquad (2)$$

From (1)

$$H\left(X/_Y\right) = H(X,Y) - H(Y) \qquad (3)$$

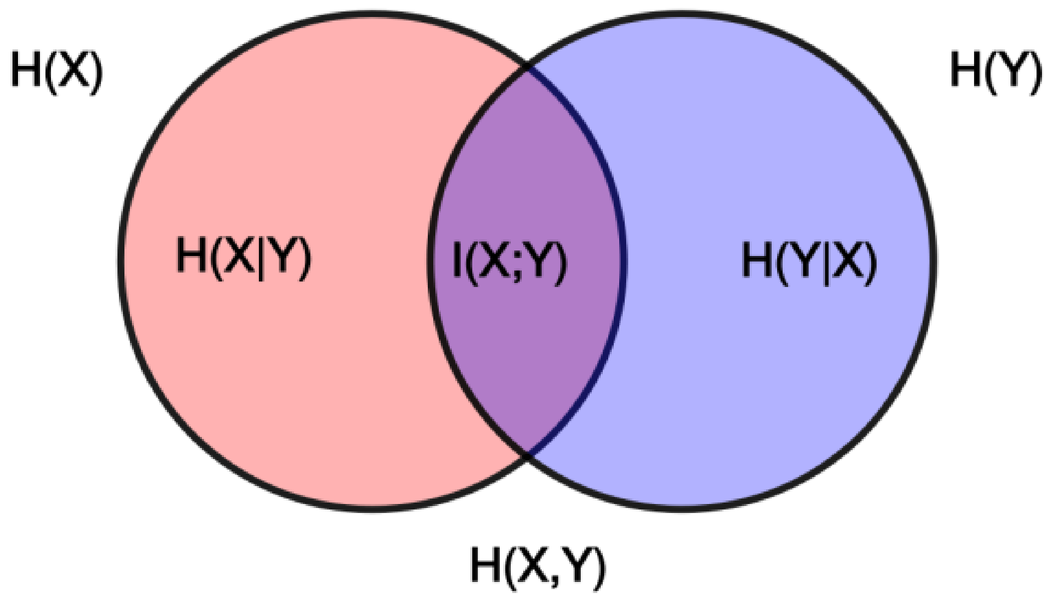
Substitute eqn(3) in (2)

$$I(X;Y) = H(X) - \left(H(X,Y) - H(Y)\right)$$

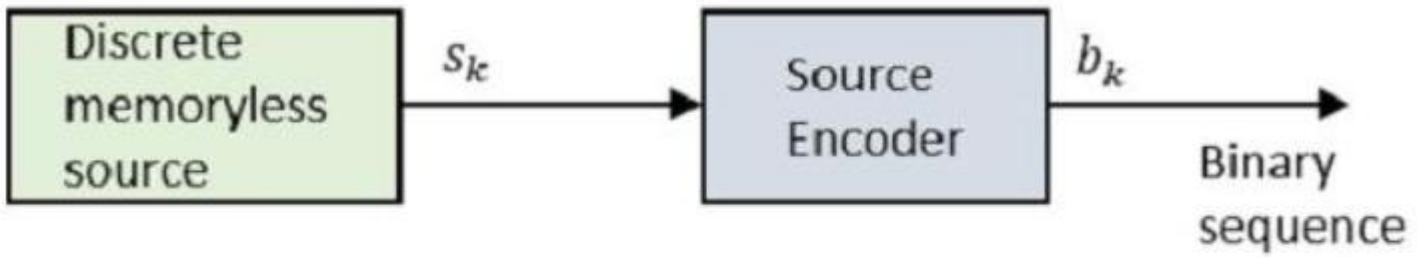
$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Hence proved

Venn diagram showing additive and subtractive relationships various information measures associated with correlated variables X and Y



Source Coding



- The Code produced by a discrete memoryless source, has to be efficiently represented.
- For this there are code words, which represent these source codes.
- Source encoding is the process by which the output of an information source is converted to *r-ary* sequence, where *r* is the number of different symbols used in this transformation process.

- S_k is the output of the discrete memoryless source and b_k is the output of the source encoder which is represented by 0s and 1s.
- The encoded sequence is such that it is conveniently decoded at the receiver.
- Let the source alphabet ‘S’ consist of ‘q’ number of source message symbols given by

$$S = \{s_1, s_2, \dots \dots .\, s_q\}$$

- Let another alphabet ‘X’ called the code alphabet consist of ‘r’ number of coding symbols given by

$$X = \{x_1, x_2, \dots \dots .\, x_r\}$$

- The term coding can be defined as transformation of the source symbols into some sequence of symbols from code alphabet X.

Example

- Consider a source S emitting only four symbols which are to be encoded with binary coding.

Then

$$S = \{s_1, s_2, s_3, s_4\}$$

and

$$X = \{0,1\}$$

The code words with a pair of code symbols are

00, 01,10,11

Different Types of Source Codes

- **Non-Singular Code**

  A code is said to be non singular if and only if all the code words are distinct and easily distinguishable from one another.

- **Uniquely Decodable Code**

  A non singular code is said to be Uniquely Decodable if every code word present in a long received sequence can be uniquely identified.

INSTANTANEOUS CODES OR PREFIX CODES

- The sufficient condition for Instantaneous Code or Prefix Code is that no complete code word be a prefix of any other code word.
- It should be uniquely decodable code.
- The necessary and sufficient condition for Instantaneous Code or Prefix Code is it should satisfy Kraft-McMillan inequality which is given by

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

- where l_k is the length of k^{th} code word.

- Example

$S = \{s_1 \; s_2 \; s_3 \; s_4\}$ $\quad X = \{0, 1\}$

	I	II	III	IV
s_1	0	0	10	0
s_2	0	010	00	10
s_3	0	01	11	110
s_4	0	10	110	111

Instantaneous

Singular

Non singular
Not uniquely decodable

$010 \rightarrow s_2$
$s_1 s_4$
$s_3 s_1$

Uniquely decodable.
But not instantaneous

Another Example

Table 1: Illustrating the definition of prefix code

Symbol	Prob.of Occurrence	Code I	Code II	Code III
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

Non singular
Not uniquely decodable

Instantaneous

Uniquely decodable.
But not instantaneous

CONSTRUCTION OF AN INSTANTANEOUS CODE

Example

Construct an instantaneous code corresponding to the source alphabet S={s_1 s_2 s_3 s_4 s_5} which is mapped to the code alphabet x={0,1}.

- We might start by assigning 0 to symbol s_1

$$s_1 \to 0$$

- If this is the case, then all other source symbols *must* correspond to code words beginning with 1.
- We cannot let correspond to the single symbol code word 1; this would leave us with no symbols with which to start the remaining three code words.

$$s_2 \rightarrow 10$$

- This, in turn, would require us to start the remaining code words with 11.

$$s_3 \rightarrow 110$$

- Then the only three-digit prefix still unused is 111 and we might set

$$s_4 \rightarrow 1110$$

- and last possibility is

$$s_5 \rightarrow 1111$$

- After constructed the code check for Kraft-McMillan inequality

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

i.e., $2^{-1}+2^{-2}+2^{-3}+2^{-4}+2^{-4} \leq 1$

Alternate code set

- we were to select a 2-digit code word to represent s_1,

$$s_1 \to 00$$

- Then we can set

$$s_2 \to 01$$

- we still have two prefixes of length 2 which are unused.

$$s_3 \to 10$$

$$s_4 \to 110$$

$$s_5 \to 111$$

Code Efficiency and Redundancy

- The average code length(average number of bits per symbol) of the source is defined as

$$\bar{L} = \sum_{k=0}^{K-1} P_k l_k$$

Let L_{min} denote the minimum possible value of $\bar{L}$.

- Then define the **Coding Efficiency** of the source encoder as

$$\eta = \frac{L_{min}}{\bar{L}} = \frac{H(S)}{\bar{L}}$$

Coding Redundancy is denoted by R_η which is defined as

$$R_\eta = 1 - \eta$$

Q: In the above prefix coding example assume that the source symbols have probabilities

$p_1 = 1/2$, $p_2 = 1/6$, $p_3 = 1/6$, $p_4 = 1/9$, $p_5 = 1/18$

Find the efficiency and redundancy of both code word sets.

NOISELESS CODING THEOREM (BOUNDS ON OPTIMAL CODE LENGTH)

- It states that for a discrete memoryless source (DMS), with entropy *H(S)*, the average code word length per symbol is bounded as

$$H(s) \leq \bar{L} \leq 1 + H(s)$$

- This theorem is also called Shannon's first theorem or fundamental theorem or Shannon's coding theorem.

Proof

- Suppose that the code alphabets consist of r symbols and the source alphabet consist of q symbols.
- By Shannon, the length of any code word l_i can be known using the formula,

$$l_i = log_r \frac{1}{P_i} \qquad (1)$$

- If l_i happens to be a fraction, then it is rounded off to next integer as given by

$$log_r \frac{1}{P_i} \leq l_i \leq 1 + log_r \frac{1}{P_i} \qquad (2)$$

- Using the property of logarithms, eqn (2) can be written as

$$\frac{log_2 \frac{1}{P_i}}{log_2 r} \leq l_i \leq 1 + \frac{log_2 \frac{1}{P_i}}{log_2 r} \qquad (3)$$

- Multiplying throughout by p_i and taking summation for all i varying from 1 to q, we get

$$\frac{1}{log_2 r}\sum_{i=1}^{q} P_i \, log_2 \frac{1}{P_i} \leq \sum_{i=1}^{q} P_i \, l_i \leq \sum_{i=1}^{q} P_i + \frac{1}{log_2 r}\sum_{i=1}^{q} P_i \, log_2 \frac{1}{P_i} \qquad (4)$$

- We know,

$$\sum_{i=1}^{q} P_i \, log_2 \frac{1}{P_i} = H(s) \qquad \sum_{i=1}^{q} P_i = 1 \qquad \sum_{i=1}^{q} P_i \, l_i = \bar{L}$$

- Substituting the above equations in eqn(4), we get

$$\frac{1}{log_2 r} H(s) \leq \bar{L} \leq 1 + \frac{1}{log_2 r} H(s)$$

- If the code alphabet consist of 2 symbols (i.e, binary coding), then the above equation becomes

$$H(s) \leq \bar{L} \leq 1 + H(s)$$

$$\bar{L} \geq H(S)$$

CONSTRUCTION OF BASIC SOURCE CODES

Shannon – Fano Algorithm

- An efficient code can be obtained

Steps for Algorithm

1. The symbols are arranged according to decreasing probabilities.
2. The symbols are divided into two sets so that the sum of probabilities in each set is approximately equal.
3. Assign ‘0’ to the upper set and ‘1’ to the lower set.
4. Continue this process, each time partitioning the sets with almost equal probabilities as possible until further partitioning is not possible.

Example

- An information source produces sequences of independent symbols x_1, x_2, x_3, x_4, x_5, x_6 with corresponding probabilities 0.30, 0.25, 0.20, 0.12, 0.08, 0.05. Construct a binary code using Shannon-Fano coding procedure. Also find its efficiency and redundancy.

Solution:

Symbol	Probability	Step I	Step II	Step III	Step IV	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_5	0.08	1	1	1	0	1110
x_6	0.05	1	1	1	1	1111

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

$$= 2.36 bits/symbol$$

Symbol	Code	Code word length
x_1	00	2
x_2	01	2
x_3	10	2
x_4	110	3
x_5	1110	4
x_6	1111	4

$$\bar{L} = \sum_{i=1}^{6} P(x_i)\, l_i$$

$$= 2.38\ bits/symbol$$

$$Efficiency, \eta = \frac{H(X)}{\bar{L}} \quad = 0.99 = 99\%$$

$$Redundancy, R_\eta = 1 - \eta \quad = 0.01 = 1\%$$

Example

- An information source produces sequences of independent symbols x_1, x_2, x_3, x_4, x_5, x_6 with corresponding probabilities 0.4, 0.1, 0.2, 0.2, 0.07, 0.03. Construct a binary code using Shannon-Fano coding procedure. Also find its efficiency and redundancy.

Symbol	Probability	Step I	Step II	Step III	Step IV	Code
x_1	0.4	0	0			00
x_3	0.2	0	1			01
x_4	0.2	1	0			10
x_2	0.1	1	1	0		110
x_5	0.07	1	1	1	0	1110
x_6	0.03	1	1	1	1	1111

Symbol	Code	Code word length
x_1	00	2
x_2	110	3
x_3	01	2
x_4	10	2
x_5	1110	4
x_6	1111	4

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

$$= 2.21 bits/symbol$$

$$\bar{L} = \sum_{i=1}^{6} P(x_i) l_i$$

$$= 2.3\ bits/symbol$$

$$Efficiency, \eta = \frac{H(X)}{\bar{L}}$$

$$= 0.9608 = 96.08\%$$

$$Redundancy, R_\eta = 1 - \eta \quad = 0.0392 = 3.92\%$$

Huffman Coding

- Better than Shannon-Fano coding
- For a source of given entropy, this code gives minimum average word length with least redundancy.
- It is called as optimum code.

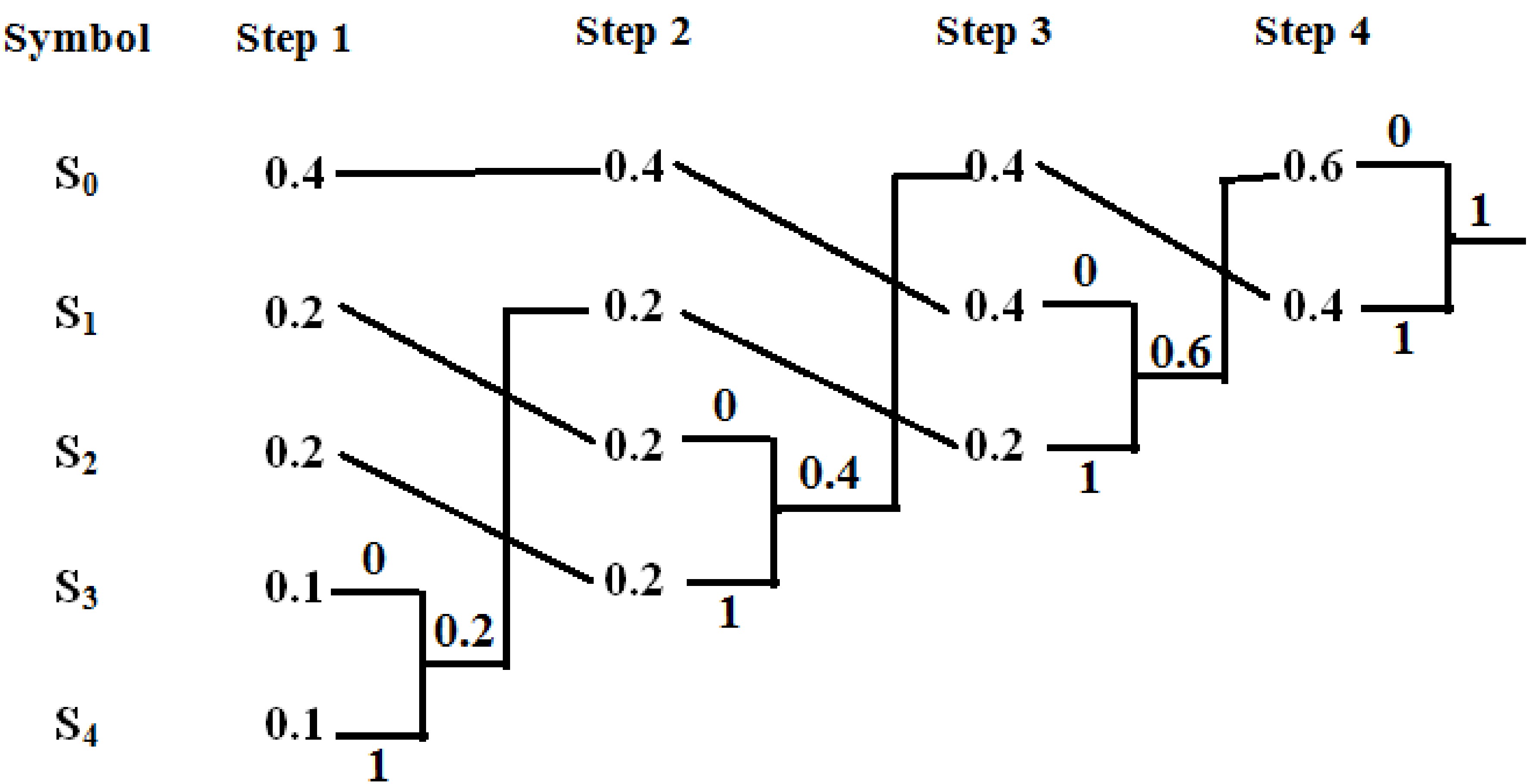
Algorithm

- Symbols are arranged in decreasing order of probability.
- Combine the two messages of lowest probability. Write their combined probability. Assign ‘0’ in the upper path and ‘1’ in the lower path.
- The sum can be placed anywhere within the group in such a way that all the probabilities are in decreasing order.
- Repeat the reduction step, ie, combine next two messages of lowest probabilities and continue this process till only two probabilities remain.
- The code for each source symbol is found by working backward and tracing the sequence of 0’s and 1’s assigned to that symbol as well as its successors. Or the code word can be found by reversing the code word obtained in the forward path.

Example 1

- Find the Huffman code for the following symbols.

Symbol	Probabilities
S_0	0.4
S_1	0.2
S_2	0.2
S_3	0.1
S_4	0.1



Symbol
Step 1
Step 2
Step 3
Step 4
S_0
S_1
S_2
S_3
S_4
0.4
0.2
0.2
0.1
0.1
0
1
0.2
0.4
0.2
0.2
0.2
0
1
0.4
0.4
0.4
0.2
0
1
0.6
0.6
0.4
0
1

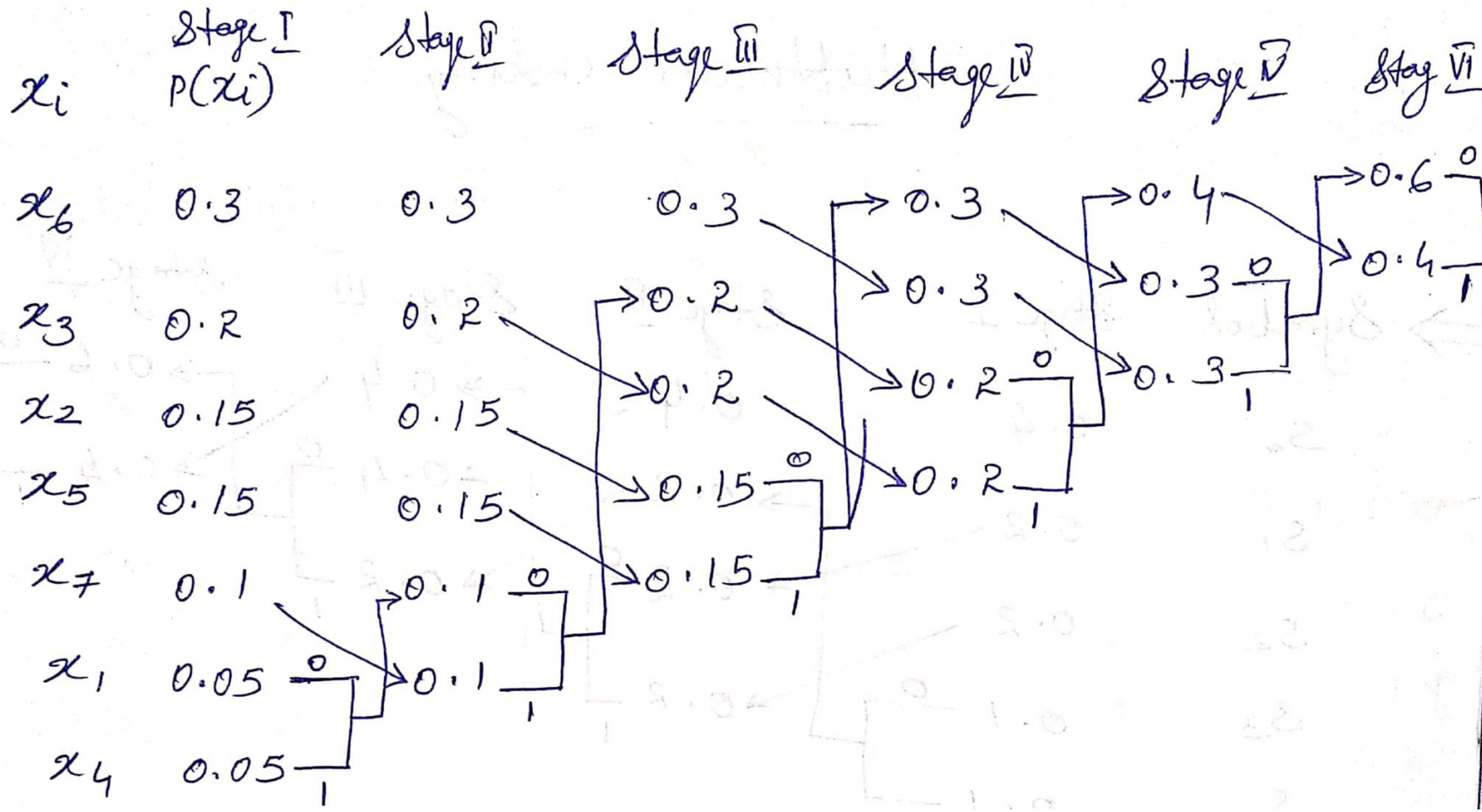
Symbol	Probability	Codeword in forward path	Codeword (By reversing Codeword in forward path)
S_0	0.4	00	00
S_1	0.2	01	10
S_2	0.2	11	11
S_3	0.1	010	010
S_4	0.1	110	011

Example 2 (Previous Univ Question)

- Determine the Huffman code for the following messages with their probabilities given

Symbol	Probabilities
x_1	0.05
x_2	0.15
x_3	0.2
x_4	0.05
X_5	0.15
X_6	0.3
X_7	0.1

- Also find its efficiency and redundancy.


Stage I
Stage II
Stage III
Stage IV
Stage V
Stage VI
xi
P(xi)
x6 0.3 0.3 0.3 0.3 0.4 0.6
x3 0.2 0.2 0.2 0.3 0.3 0.4
x2 0.15 0.15 0.2 0.2 0.3
x5 0.15 0.15 0.15 0.2
x7 0.1 0.1 0.15
x1 0.05 0.1
x4 0.05
0
1

Symbol	Probabilities	Codeword in forward path	Codeword (By reversing Codeword in forward path)
x_1	0.05	0001	1000
x_2	0.15	000	000
x_3	0.2	11	11
x_4	0.05	1001	1001
X_5	0.15	100	001
X_6	0.3	10	01
X_7	0.1	101	101

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

$= 2.57$ bits/symbol

$$\bar{L} = \sum_{i=1}^{6} P(x_i)\, l_i$$

$= 2.6$ bits

$$Efficiency, \eta = \frac{H(X)}{\bar{L}}$$

$= 98.85\%$

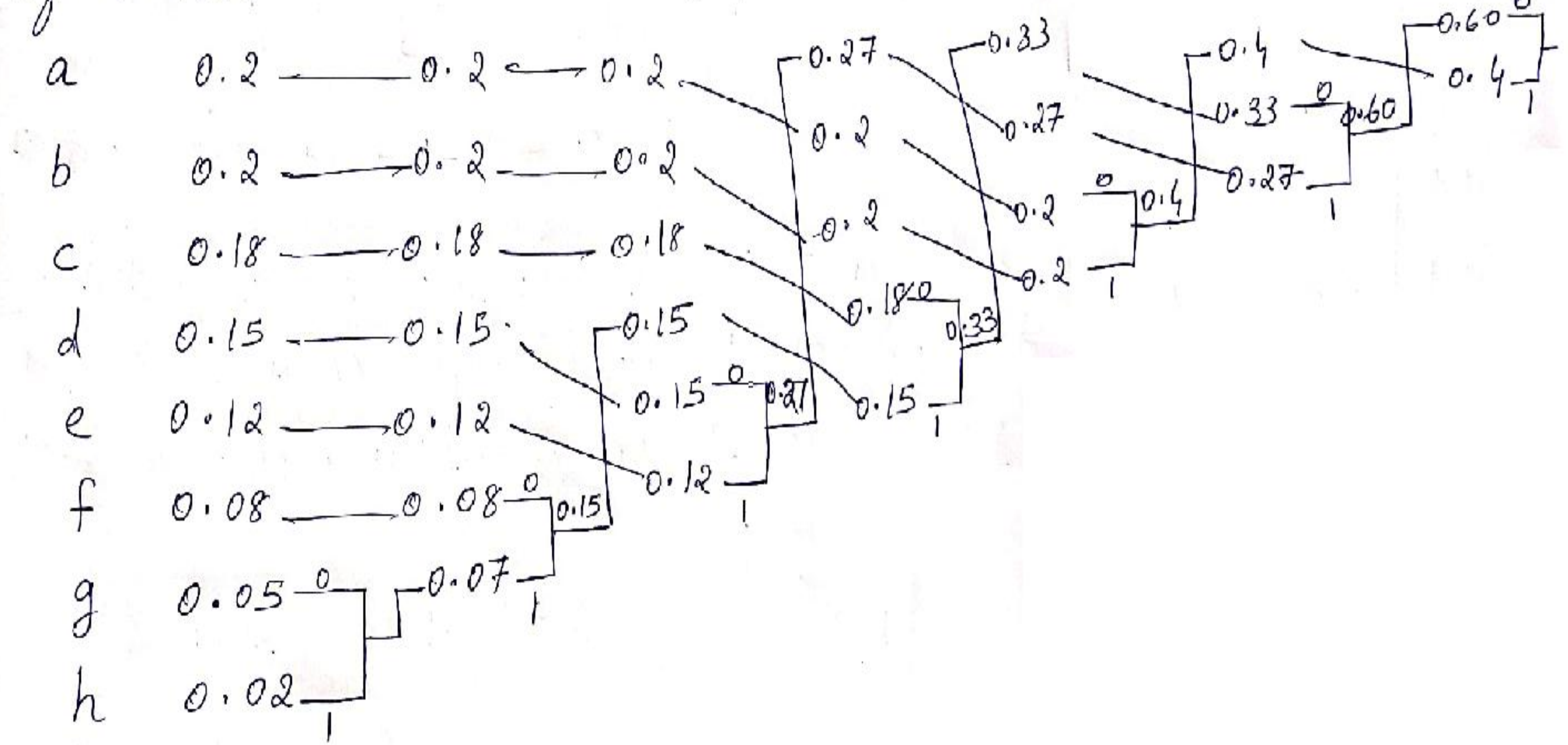
$$Redundancy, R_\eta = 1 - \eta$$

$= 1.15\%$

Example 3 (Previous Univ Question)

- Consider a source with 8 alphabets, a to h with respective probabilities 0.2, 0.2, 0.18, 0.15, 0.12, 0.08, 0.05 and 0.02. Construct a minimum redundancy code and determine the code efficiency.

Symbol Prob.



	Forward path	Actual code word
a	01	10
b	11	11
c	000	000
d	010	010
e	110	011
f	0100	0010
g	01100	00110
h	11100	00111

Example 4 (Previous Univ Question)

- An information source produces sequences of independent symbols A, B, C, D, E, F, G with corresponding probabilities 1/3, 1/27, 1/3, 1/9, 1/9, 1/27, 1/27. Construct a binary code and determine its efficiency and redundancy using

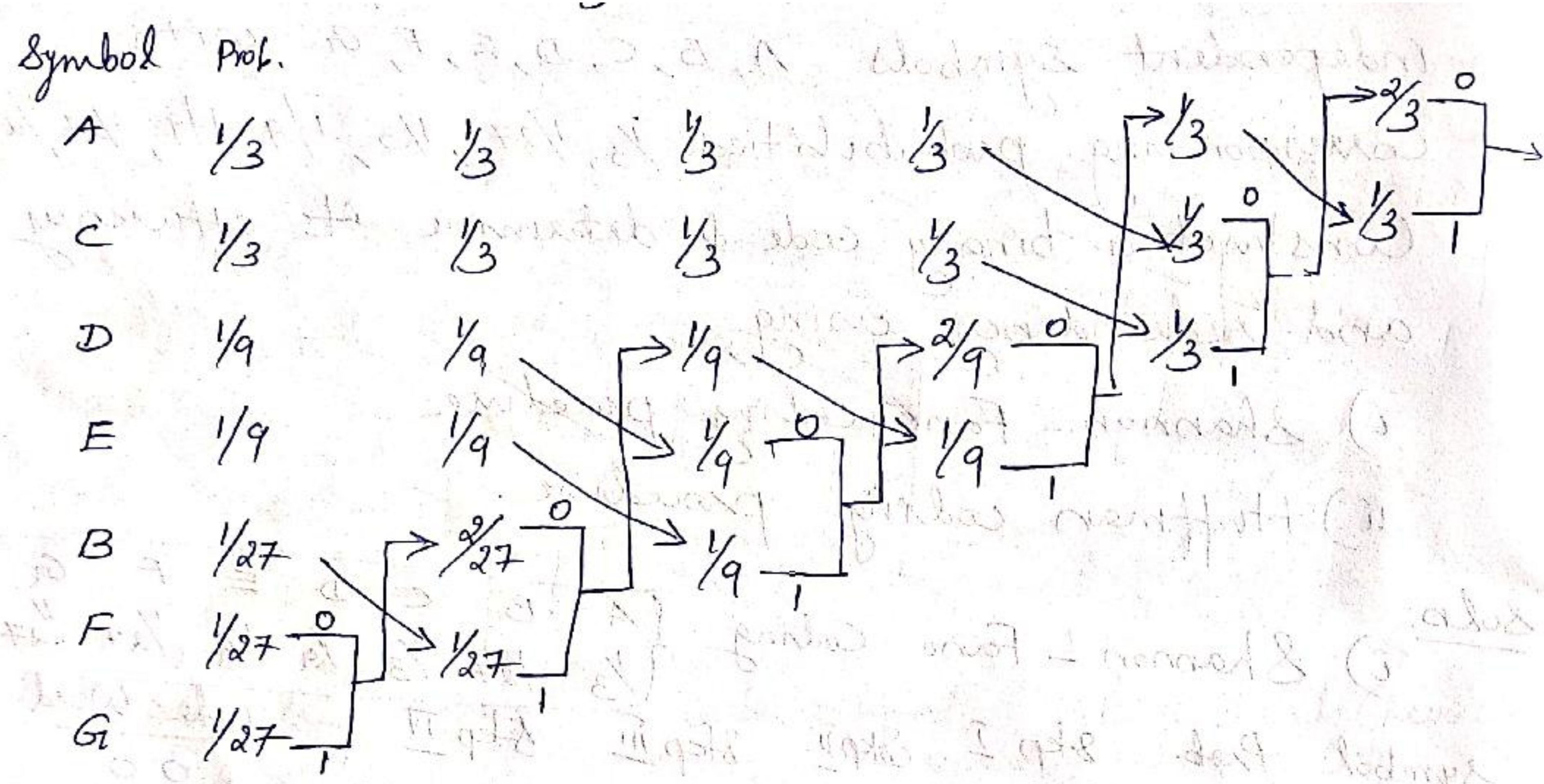
  i) Shannon-Fanocodingprocedure

  ii) Huffrnan coding procedure.

Shanon-Fano coding

Symbol	Prob.	Step I	Step II	Step III	Step IV	Code wor
A	$1/3$	0	0			0 0
C	$1/3$	0	1			0 1
D	$1/9$	1	0	0		1 0 0
E	$1/9$	1	0	1		1 0 1
B	$1/27$	1	1	0		1 1 0
F	$1/27$	1	1	1	0	1 1 1 0
G	$1/27$	1	1	1	1	1 1 1 1

Huffman Coding



Symbol	Prob.	Code in Forward	Code Word (By Reversing)
A	1/3	0 0	0 0
B	1/27	1 1 1	1 1 1
C	1/3	1 0	0 1
D	1/9	0 0 1	1 0 0
E	1/9	1 0 1	1 0 1
F	1/27	0 0 1 1	1 1 0 0
G	1/27	1 0 1 1	1 1 0 1