# Indian Institute of Information Technology, Vadodara

**Course** - Information Theory, Inference and Learning
**Course Code** - SC307
**Prof. P Shah**

MidSem

---



1. (a) What is the minimum value of $\mathrm{H}(p) = \mathrm{H}(p_1, p_2, p_3, \ldots, p_n)$ as $p$ ranges over the set of $n$ dimensional probability vectors? Find all $p$'s that achieve this minimum.

   (b) Given an ensemble $\chi = \{x, A_x, P_x\}$, with $A_x = \{a, e, i, o, u\}$ and corresponding $P_x = \{0.22, 0.34, 0.17, 0.19, 0.8\}$.

      i. Give an expression for the number of bits of information received when we learn that a particular vowel is either i or u.

      ii. Assume that a programmer comes up with an encoding algorithm such that in his code the expected code length in bits of an encoded message transmitting 100 vowels is 197 bits. How good or bad will this code be? Explain your claim.

   (c) A dog walks on the integers, possibly reversing direction(positive and negative directions denoted by $\{+, -\}$) at each step with probability $p = 0.1$. Let the initial direction of the dog be $X_0 = +$. Find $H(X_0, X_1, X_2, \ldots, X_n)$ for any n.

2. Shannon is working on information theory with a landmark paper, *"A Mathematical Theory of Communication"*. He is constructing a prefix code based on a set of characters and their frequencies. He is trying to achieve the lowest possible expected code word length using three symbols $\{0, 1, 2\}$. Below are the characters with their associated frequencies.

| Character | Frequency |
|---|---|
| S | 13 |
| H | 45 |
| I | 18 |
| A | 53 |
| N | 22 |
| O | 45 |
| L | 66 |

   Shannon has developed Alzheimer's disease, so he needs your help to create marvels of the trits(similar to bits) revolution.

   (a) Find the codeword for each character using Huffman coding tree.

   (b) Using the above compression scheme write the trit sequence that encodes HAILSHANNON.

   (c) What is the expected number of trits needed to be transmitted fore the data coming from the source, i.e. the characters and the relative frequencies described above?

3. Answer the following question:

   (a) Consider an ensemble $\chi := \{x, A_x, P_x\}$, where the alphabets are $\{0, 1, 2\}$ and $P(x = 0) = 0.2; P(x = 1) = P(x = 2) = 0.4$. Encode the sequence X = 2, 1, 0 using arithmetic coding scheme.

   (b) Discuss how arithmetic coding scheme rectifies problems with Huffman's coding scheme.

(c) Briefly discuss a way to generate random string of desired length using arithmetic coding concept. Also comment on why this string can't be of arbitrary length when working with computers.

4. Consider a channel with transition matrix:

$$Q = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Here the entry in the $x^{th}$ row and the $y^{th}$ column denotes the conditional probability $p(\gamma = y|X = x)$ that y is received when x is sent. What is the capacity of this channel? For what input probability distribution is this capacity achieved?

5. Let us assume that three symbols a, b, c used for generating codewords are physically represented by voltage waveforms off different durations, i.e. $T_a = 2$, $T_b = 1$ and $T_c = 1$ time units. Generation of codewords essentially amounts to creation of different waveforms generated by concatenating the symbol waveforms physically. Let N(T) denote the number of allowed waveforms of duration T time units. Calculate

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T}$$

6. Fun Problem: You meet Fred. Fred tells you he has two brothers, Alf and Bob.

   (a) What is the probability that Fred is older that Bob?

   (b) Fred tells you that he is older than Alf. Now, what is the probability that Fred is older than Bob?

---

"Forgetting used to be a failing, a waste, a sign of senility. Now it takes effort. It may be as important as remembering."

*James Gleick(The Information: A History, a Theory, a Flood)*

# Answers