Information Theory

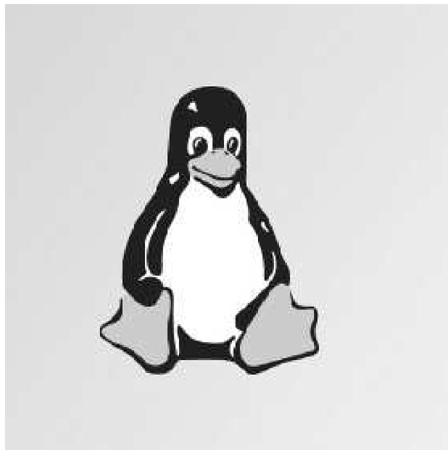
1. In this problem we aim to calculate the average information content in English language.

(a) Calculate the average information in bits/character for English language assuming that each of the 26 characters in the alphabets occur with equal probability and we neglect spaces and punctuations.

(b) Since the alphabetic characters do not appear with equal frequency in the English language (or any other language), the answer to part (a) will represent an upper bound on average information content per character. Repeat (a) under the assumption that alphabetic characters occur with following probabilities

p=0.1	for the letters a,e,o,t
p=0.07	for the letters h,i,n,r,s
p=0.02	for the letters c,d,f,l,m,p,u,y
p=0.01	for the letters b,g,j,k,q,v,w,x,z

2. For this problem we consider source-coding of the following image. Assume that the image has 256×256 pixels. It is observed that image has only four shades i.e. (Black, dark gray,light gray and white).



(a) Calculate the entropy of a pixel assuming all shades are equally likely.

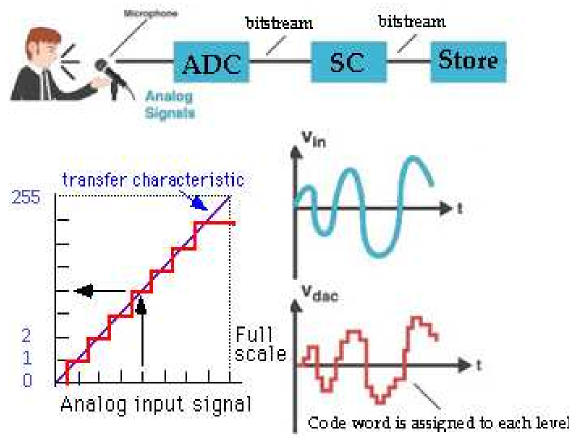
Now assume that

$$
\begin{aligned}
P(x = LG) &= 0.5 \\
P(x = DG) &= 0.10 \\
P(x = W) &= 0.15 \\
P(x = B) &= 0.25
\end{aligned}
$$

(b) What would be entropy of a pixel and image in this case?

(c) How much information redundancy is removed in this scenario?

3. For this problem we consider source coding of audio signal. The audio signal is converted into digital symbols. The overall process is illustrated below:



Assuming that audio signal is sampled at 8000 samples/second and the AD convertor can produce a total of 256 possible symbols.

(a) Calculate the entropy of each sample assuming all samples are equally likely.

(b) What would be the source rate R of this system?

(c) Now assume that most of the audio content consists of silence, therefore symbols $1 \cdots 10$ occur with joint probability of 40%, While symbols $11 \cdots 200$ occur with probability of 50% and symbols $201 \cdots 256$ occur with probability of 10%. Calculate the entropy on this case.

(d) What would be the source rate R in this scenario.

(e) How much information redundancy reduced through this scheme.

4. In this problem we explore the dynamics of Shannon's channel capacity theorem. For this problem consider that we have 4000 Hz of bandwidth available.

(a) What would be the achievable data rate if available SNR is 10dB.

(b) What should be the minimal SNR if data rate of 10 Kbits/s are to be achieved.

(c) What would be the bandwidth efficiency in this case.

5. A telephone channel has a bandwidth $W = 3000$ Hz and signal to noise ratio $P_{av}\ W/N_0$ of 400 (26 dB). Suppose we characterize the channel as a band-limited AWGN waveform channel then

(a) Determine the capacity of the channel in bits/s.

(b) Is this capacity sufficient to support the transmission of speech signal that has been sampled and quantized through log-PCM.

(c) Usually, channel impairments other than additive noise limit the transmission over telephone channel considered in (a.). Suppose that transmission rate of 0.7C is achievable in practice without channel encoding, which speech source coding techniques can be used to sufficiently compress the data to fit in the bandwidth available.

6. Assume a DMS source produces 6 symbols $x = \{a, b, c, d, e, f, g\}$ with following probabilities $p(x) = \{0.25, 0.2, 0.15, 0.15, 0.1, 0.1, 0.05\}$.

(a) Calculate the entropy H of the source.

(b) Construct the Shannon-Fano code for this system.

(c) Construct the Huffman code for this system.

(d) which system is efficient and why?