20SK: Exercise #5

Source encoding

Jan Přikryl

December 12, 2017

Introduction

Source encoding: *Remove rendunant information.*

Medium: *Reduce message entropy to approach* $H(X)$.

Character-based methods (*entropy encoding*):

- Huffman coding
- Arithmetic coding
- Asymetric Numeral Systems (ANS)

Dictionary-based methods:

- LZ77, LZW
- LZMA, Deflate, etc.

Arithmetic Coding

Nearly optimal Entropy encoding

Huffman code optimal for $p_i = 1/2^k$.

Encode the whole message into a binary fraction:

1. resulting number $n \in [0, 1)$,
2. represented using *arbitrary precision interval arithmetic*.

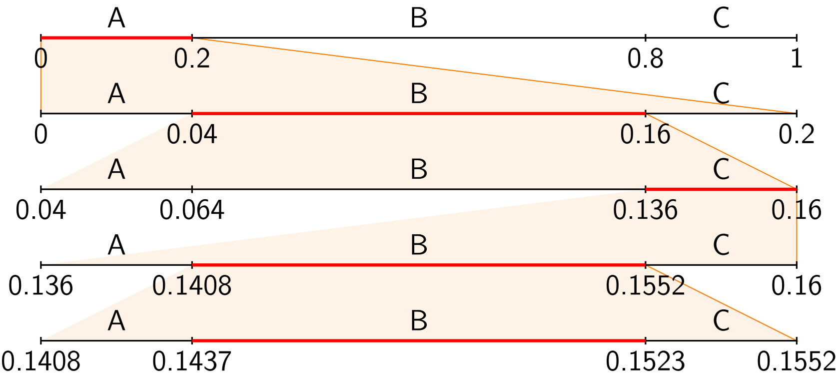
Principle:

1. stochastic model of message data using symbol probabilities, $\sum_i p_i = 1$,
2. *recursive subdivision* of existing sub-intervals.

Arithmetic Coding

Encoding process

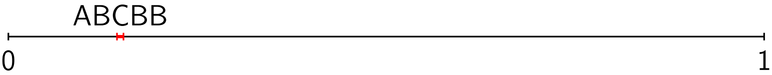
Encode sequence "ABCBB" with $p(A) = 0.2$, $p(B) = 0.6$, $p(C) = 0.2$:



Arithmetic Coding

Resulting message

Encoded sequence "ABCBB" corresponds to $n \in (0.1437, 0.1523)$ in binary form:



Such n is for example $(0.0010011)_b$, i.e.

$$n = \frac{1}{8} + \frac{1}{64} + \frac{1}{128} = 0.1484.$$

Problem 1

1. Create function `binstr=aenc(text,probs)` generating arithmetic code for sequence in vector `text` with symbol probabilities in `probs`.
   1.1 compute cumulative probabilities for `probs`
   1.2 encode the sequence into an floating point interval for the number n
   1.3 find a representaion of n as a sum binary fractions

Problem 1 continued

4. Create function `[eps,ber]=bsc_ber(n)` that will simulate the BER of BSC for packets of length `n` and different values of transition probability.
   Note 1: BER stands for *bit error ratio*, given as a ratio of number of bit errors to the total number of bits transferred.
   Note 2: BER may also denote *bit error rate*, the number of bit errors per unit time.
   4.1 Use `linspace` to define equally spaced crossover probability values ranging from 0.01 to 0.5.
   4.2 Use `bsc` from previous task to compute BSC output for given eps
   4.3 Comparing `x~=y` gives an vector with 1 at places where error occured
   4.4 Summing this vector gives you the number of bit errors
5. Plot the graph ber $= \mathrm{bsc}(\epsilon)$.

Additive White Gaussian Noise Channel

Review

The channel is bandlimited to $[-W, W]$, the noise is Gaussian and white with power spectral density $N_0/2$ (two-sided). Channel input satisfies power constraint P.

Definition (Capacity of AWGN)

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right) = W \log_2 \left(1 + \frac{S}{N}\right),$$

where S/N is the signal-to-noise power ratio (watts, volts, not decibels).

Definition (Q-function)

A probability of error exceeding x is given as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$$

Problem 2

Binary data are transmitted over AWGN with noise power spectral density $N_0/2$ using BPSK with energy E and hard-decision decoding. In this case the channel can be modelled as BSC.

1. Plot the error probability of the channel as a function of

$$\gamma = \frac{E}{N_0}, \qquad \gamma \in [-20\,\mathrm{dB}, 20\,\mathrm{dB}].$$

   1.1 Create vector `gamma_db=-20:0.1:20` and covert it from decibels.
   1.2 The error probability of BPSK with optimal detection is $p = Q(\sqrt{2\gamma})$.
   1.3 Plot it using `semilogx(gamma,p)`.

Problem 2 continued

2. Plot the capacity of the channel as a function of γ.
   2.1 Use the `gamma` from previous task.
   2.2 The capacity of BSC is given by the binary entropy of the channel, which again depends on error probability computed in previous task.
3. What seems to be the limit on error probability? Explain.

Problem 3

1. Plot the capacity of AWGN channel with bandwidth $W = 3000\,\text{Hz}$ as a function of P/N_0 for values of P/N_0 between -20 and $30\,\text{dB}$.
   1.1 adopt the approach we used for `gamma`, i.e. `pn0_db=[-20:0.1:30];` and so on
   1.2 use the standard AWGN capacity formula
2. Now fix the P/N_0 at $25\,\text{dB}$ and plot the capacity as a function of W in the range from 10 to $10000\,\text{Hz}$.
3. Study the limit cases of capacity. Is there any difference between channel behaviour dependent on noise and bandwidth?
4. Open another figure and plot three bandwidth-capacity graphs for $P/N_0 = 25, 40$ and $55\,\text{dB}$. What do you observe?