Entropy Coding

Task

We have an 8-symbol source with the following discrete probability distribution.

Table 1 Probability distribution of the 8-symbol source

Symbol	a	b	c	d	e	f	g	h
Probability	0.31	0.12	0.06	0.2	0.01	0.18	0.02	0.10

Please investigate entropy coding for this source.

Basic Requirements

- Implement a sequence generator with specified statistics in Table 1.
  - Generate five sequences with different lengths: 500, 1000, 5000, 10000, 20000, 30000, 50000, 100000;
  - Estimate the distributions for these sequence, and analyzes the estimation accuracy against the ideal probability distribution with respect to sequence lengths;
  - Estimate sample entropies for these sequence, and analyzes the estimation accuracy against the ideal source entropy with respect to sequence lengths;
- Construct the Huffman coding tree with the source distribution in Table 1, code each sequence and calculate the average code length.
- Code each sequence by arithmetic coding with the source distribution in Table 1, and calculate the average code length.
- Compare the coding efficiency of Huffman coding and arithmetic coding with the theoretic bound given by entropy.

Materials

Tools

- **CODEC for arithmetic coding:** FastAC.zip