

Boris Kudryashov

ITMO University

October 19, 2016

- 1 Build Huffman code for source with letters probabilities equal to (0,3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05). Compare the average length of code words with entropy of the source.

- 1 Build Huffman code for source with letters probabilities equal to (0,3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05). Compare the average length of code words with entropy of the source.
- 2 Consider ternary constant source X with the following letters probabilities:
 - ① $p(a) = 1/2, p(b) = 1/4, p(c) = 1/4$;
 - ② $p(a) = p(b) = p(c) = 1/3$.

Build Huffman codes for letters of X , X^2 , X^3 .
Compare code rate and the entropy of the source.

- 3 Consider the dependency between Huffman code rate and the length of correlated blocks for binary constant source with probability of 1 equal to 0.1.

- 4 *Optimal coding of equiprobable letters* Build Huffman code for source, which chooses letters from alphabet of size M with equal probabilities. Calculate the average length of code words and redundancy like function of M , calculate upper and lower bounds of redundancy.

4 (hint) *Intermediate steps* Huffman code for this source contains $D = 2 \cdot 2^{\lfloor \log M \rfloor} - M$ words of length $\lfloor \log M \rfloor$ and $M - D$ words of length $\lfloor \log M \rfloor + 1$. Average length of code words and redundancy are equal to

$$\bar{l} = \lfloor \log M \rfloor + 2 - \frac{2}{M} 2^{\lfloor \log M \rfloor},$$

$$r = \bar{l} - \log M = 2 - d - 2 \cdot 2^{-d},$$

where $d = \log M - \lfloor \log M \rfloor$, $d \in [0, 1)$. Differentiation with respect to d shows, that maximum of redundancy is achieved when $d = 1 - \log \log e$ and is equal to $1 + \log \log e - \log e \approx 0,0861$.

- 5 Ternary source is specified by Markov chain with the following transition matrix:

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$

Calculate $H(X)$, $H_n(X)$, $n = 1, 2, \dots$, $H(X/X^n)$, $n = 1, 2, \dots$, assuming that initial distribution on letters is stationary distribution. Find code rate, when coding two ensembles X and X^2 with Huffman coding. Provide a method for encoding, when code rate is equal to the speed of information creation by the source.

- 6 *Encoding of length of sequences with non-uniform code.*
 Sequence at the output of binary source is splitted in the only way to the subsequences of a random length like $1, 01, \dots, 0^{L-1}1, 0^L$. Huffman code is used to encode such sequences. For binary constant source with probability of 1 equal to $p=0,1$, consider the dependence between code rate and the value L .
 Code rate in this case is a value $R = \bar{n}/\tau$, where τ – average length of sequence to be encoded; \bar{n} – average length of code words.

- 7 *Non-uniform encoding by input and output (VV). Tunstall code.*
- Recall, that the redundancy of the Huffman code is determined to a large extent by the maximum among the probabilities of letters. Hence, when encoding the source $X = p(x)$ the letter $x_0 \in X$ should be found, which has the maximal probability. Then we should build new ensemble X_1 , which consists of all letters from $X \setminus x_0$ and of all pairs (x_0, x) , $x \in X$. Probabilities of pairs of letters are calculated like products of probabilities of corresponding letters. Such extension of the alphabet can be continued, and on each next step, the maximal probability of letters of the extended alphabet becomes smaller. Thus the redundancy of code applied to the extended alphabet also becomes smaller.
- For a random binary ensemble compare this way of encoding with coding of length of sequences, considered in previous problem.

- 8 Build Shannon code of the source from task 1. Compare average length with the average length of Huffman code codewords and with entropy of the source.

- 8 Build Shannon code of the source from task 1. Compare average length with the average length of Huffman code codewords and with entropy of the source.
- 9 Build Gilbert-Moore code of the source from task 1. Compare average length with the average length of Huffman code codewords and with entropy of the source.

- 10 Encode with arithmetic coding the sequence 01001 at the output of the binary source where probability of 1 is 0.4 Compare the length of code sequence with the total information of the sequence. Compare the length of code sequence with total length of code sequence assuming independent encoding of characters with Shannon code.

- 10 Encode with arithmetic coding the sequence 01001 at the output of the binary source where probability of 1 is 0.4 Compare the length of code sequence with the total information of the sequence. Compare the length of code sequence with total length of code sequence assuming independent encoding of characters with Shannon code.
- 11 Explain, why end of encoding in the program ?? is equivalent to conversion of the codeword of Shannon code to codeword Gilbert-Moore code.

- 12 What changes should be done to algorithms and programs of arithmetic coding and decoding, to apply them to source, described by simple Markov chain.

- 12 What changes should be done to algorithms and programs of arithmetic coding and decoding, to apply them to source, described by simple Markov chain.
- 13 Formally, in programs at the pictures ?? and ?? there are division operations, which weren't there in algorithms of arithmetic coding and decoding. Why? Is it possible to avoid divisions?