# Boris Kudryashov

# ITMO University

October 22, 2016

1 Construct a Huffman code for the source with probability distribution (0,3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05). Compare the average codeword length with the entropy.

1 Construct a Huffman code for the source with probability distribution (0,3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05). Compare the average codeword length with the entropy.

2 Consider Discrete Memory-less Source (DMS) $X = \{a, b, c\}$ with probability distribution:
   1 $p(a) = 1/2, p(b) = 1/4, p(c) = 1/4$;
   2 $p(a) = p(b) = p(c) = 1/3$.

Find Huffman codes for $X$, $X^2$, $X^3$. Compare code rate with the entropy of the source.

3 Analyse behaviour of the Huffman code rate as a function of length $n$ of encoded blocks for the Binary Memory-less Source (BMS) with probability of 1 equal to 0.1.

4 *Optimal coding uniform distributions* Find a Huffman code for source, which chooses letters from alphabet of size $M$ with uniform distribution. Calculate the average length of code words and redundancy as a 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 codewords 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 Consider Markov chain with the probability 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 $H(X)$, $H_n(X)$, $n = 1, 2, ...$, $H(X/X^n)$, $n = 1, 2, ...$, assuming that the initial distribution on letters is the stationary distribution. Find code rate for coding ensembles $X$ and $X^2$ using Huffman coding. Suggest such a method for encoding, that code rate is equal to the information rate of the source.

6 *Non-uniform series-length coding* Sequence at the output of binary source is uniquely represented to the subsequences of a random length like $1, 01, ..., 0^{L-1}1, 0^L$. Huffman code is used to encode such sequences. For a BMS with probability of 1 equal to $p=0,1$, analyse the dependence between code rate and the value $L$ .
Code rate in this case is a value $R = \bar{n}/\tau$, where $\tau$ − average length of sequence to be encoded; $\bar{n}$ − average length of codewords.

7 *Non-uniform encoding by input and output (VV). Tunstall code.*
Recall that the redundancy of the Huffman code is determined by
the maximum probability of letters.

Thus, when encoding the source $X = p(x)$ the letter $x_0 \in X$
should be found, which has the maximal probability. Then we
should construct 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.

Compare this way of encoding to series-length coding, considered
in previous problem for a random binary ensemble

8 Construct 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 Construct 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 Construct 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 Use arithmetic coding to encode the sequence 01001 from the BMS with probability of 1 is 0.4 Compare the length of code sequence with the amount of 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  Use arithmetic coding to encode the sequence 01001 from the BMS with probability of 1 is 0.4 Compare the length of code sequence with the amount of 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 (p. 87, pic. 2.16 in the book) 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 In programs at the pictures (p. 87, pic. 2.16) and (p. 88, pic. 2.17) in the book there are division operations, which weren't there in algorithms of arithmetic coding and decoding. Why? Is it possible to avoid divisions here?