LINKÖPING UNIVERSITY

Written Exam in
Data compression
TSBK08

21st March 2022 14:00 - 18:00

Location:	FE245
Examiner:	Harald Nautsch
Teacher:	Harald Nautsch, 281361
Department:	ISY
Exam code:	TEN1
Number of problems:	7
Number of pages:	4
Permitted equipment:	Calculator, general English dictionaries
Other:	Answers can be given in English or in Swedish. The teacher will visit at 15:15 and 16:45
Grades:	0-13 U 14-19 3 20-25 4 26-30 5

1 a) Explain what an instantaneous code is.

(1 p)

b) Formulate Kraft-McMillan's inequality.

(1 p)

c) What type of coding is used in PNG images?

(1 p)

d) Explain how prediction with partial match (ppm) coding works.

(2 p)

e) Explain what a Golomb code is and what type of probability distribution it is good for.

(2 p)

f) Explain how coding and decoding of data using Burrows-Wheelers transform is done.

(2 p)

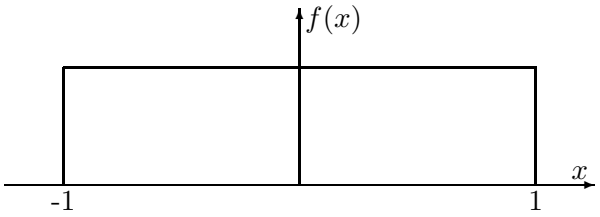
2 A stationary random source X_i is coded using an optimal code. Show that the average rate R is bounded by

$$H(X_i) \leq R < H(X_i) + 1$$

if we code one symbol with each codeword.

(4 p)

3 Let X_k be a stationary memoryless time-discrete amplitude-continuous random process. The amplitudes are uniformly distributed over the interval $[-1 \;\; 1]$.



X_k is quantized with a uniform quantizer and then source coded using an arithmetic coder. The number of symbols n that is coded with each codeword can be assumed to be large. What is the resulting rate R (in bits/symbol) as a function of the mean square error D?

(2 p)

4 A binary source has the alphabet $\mathcal{A} = \{a, \; b\}$. From a large set of test data, the probabilities of triples $p(x_i, x_{i+1}, x_{i+2})$ have been estimated as

$$\begin{array}{ll} p(a,a,a) = 63/110 & p(a,a,b) = 7/110 \\ p(a,b,a) = 4/110 & p(a,b,b) = 6/110 \\ p(b,a,a) = 7/110 & p(b,a,b) = 3/110 \\ p(b,b,a) = 6/110 & p(b,b,b) = 14/110 \end{array}$$

From these estimated probabilities we can make random models that are memoryless or Markov sources of order 1 or 2. Calculate the entropy rate for each of the three possible source models. For the two Markov models also draw the corresponding state diagrams.

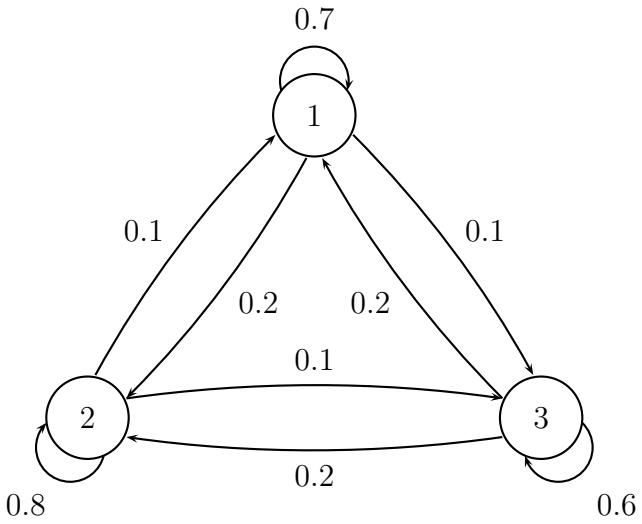
(4 p)

5 Consider the source in problem 4.

Construct a Huffman code for the triples and calculate the resulting average data rate (in bits/symbol) of the code.

(3 p)

6 A stationary Markov source X_n of order 1, with alphabet $\mathcal{A} = \{1,\ 2,\ 3\}$, is given by the state diagram below



Code the sequence

$$222311$$

using arithmetic coding. The coding should utilize the memory of the source. Give both the interval and the corresponding codeword. You can assume that the source is in state 1 when the coding starts and that all calculations are performed with infinite precision.

(4 p)

7 A source has the alphabet $\{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p\}$.

We want to code the source using LZSS.

a) Assume that we want to use a history buffer length of 256 and that we want use 4 bits to code the match lengths. What is the shortest match length that should be coded as a match instead of a sequence of single symbols?

(1 p)

b) Code the sequence beginning with

$$badbadbepppppadbepppeppo\ldots$$

Give the resulting binary codewords.

(3 p)