LINKÖPING UNIVERSITY

# Written Exam in
# **Data compression**
# **TSBK08**

20th March 2023 14:00 - 18:00

| | |
|---|---|
| **Location:** | U7,TER2 |
| **Examiner:** | Harald Nautsch |
| **Teacher:** | Harald Nautsch, 1361 |
| **Department:** | ISY |
| **Module:** | TEN1 |
| **Number of problems:** | 8 |
| **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 around 15:15 and 16:45 |
| **Grades:** | 0-13 U<br>14-19 3<br>20-25 4<br>26-30 5 |

1 a) Explain how adaptive arithmetic coding works.

(2 p)

b) Explain what universal coding is and give an example of such a coding method.

(2 p)

c) Explain what the rate-distortion function is and how it is calculated for a stationary memoryless random source.

(2 p)

2 Describe how the coding works in the following lossless image coding standards.

a) JPEG-LS

(1 p)

b) PNG

(1 p)

c) GIF

(1 p)

3 Formulate Kraft's inequality and give a proof of it.

(4 p)

4 A memoryless source has the alphabet $\mathcal{A} = \{a,\ b,\ c\}$. The symbol probabilities are

$$p(a) = 0.6,\ p(b) = 0.35,\ p(c) = 0.05$$

What is the resulting average data rate (in bits/symbol) if we code pairs of symbols from the source using a Huffman code?

(3 p)

5 A fax machine works by scanning paper documents line by line. The symbol alphabet is black and white pixels, ie $\mathcal{A} = \{b,\ w\}$. We want to make a random model $X_i$ for typical documents and calculate limits on the data rate when coding the documents.

From a large set of test documents, the following conditional probabilities $p(x_i|x_{i-1}, x_{i-2})$ (note the order) have been estimated.

$$\begin{array}{ll} p(w|w,w) = 0.9 & p(b|w,w) = 0.1 \\ p(w|w,b) = 0.85 & p(b|w,b) = 0.15 \\ p(w|b,w) = 0.3 & p(b|b,w) = 0.7 \\ p(w|b,b) = 0.2 & p(b|b,b) = 0.8 \end{array}$$

a) The given probabilities imply a Markov model of order 2. Draw the state diagram for this Markov model and calculate the stationary probabilities.

(1 p)

b) Calculate the entropies $H(X_i)$, $H(X_i|X_{i-1})$ and $H(X_i|X_{i-1}, X_{i-2})$ for the model.

(3 p)

6 Consider the source in problem 5. Use arithmetic coding to code the sequence

$$wwwbbw$$

The memory of the source should be utilized in the coder. The source can be assumed to be in state $ww$ when the coding starts. You can assume that the coder can store all probabilities and interval limits exactly. Give both the resulting interval and the codeword.

(4 p)

7 A source has the alphabet $\{a,\ b,\ c,\ d,\ e\}$. A sequence from the source is coded using LZW giving the following index sequence:

$$1, 4, 3, 1, 0, 7, 6, 11, 8, 5, 13, 15, \ldots$$

The starting dictionary is;

| index | sekvens |
|---|---|
| 0 | $a$ |
| 1 | $b$ |
| 2 | $c$ |
| 3 | $d$ |
| 4 | $e$ |

Decode the index sequence. Also give the dictionary.

(3 p)

8 A source has the alphabet $\mathcal{A} = \{a,\ b,\ c,\ d\}$. A symbol sequence of length 8 is coded using BWT and mtf. The resulting index is 6 and the mtf-coded sequence is 2,0,0,1,0,2,0,0. Decode the symbol sequence.

(3 p)