UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA, TELECOMUNICAÇÕES E INFORMÁTICA

Information and Coding (2021/22)

Problem set 1

1. Is the code $\{00, 11, 0101, 111, 1010, 100100, 0110\}$ uniquely decodable?
2. Is the (ternary) code $\{00, 012, 0110, 0112, 100, 201, 212, 22\}$ uniquely decodable?
3. Find a probability distribution $\{p_1, p_2, p_3, p_4\}$ such that there are two optimal codes that assign different lengths $\{l_1, l_2, l_3, l_4\}$ to the four symbols.
4. Explain why every compression method will expand some of its inputs, instead of compressing them.
5. Consider a Golomb code with parameter $m = 5$ (note that m is not a power of two).
   (a) According to this code, give, justifying, a sequence of bits that represents the integers $a = 12$ and $b = 13$ as efficiently as possible.
   (b) Indicate the optimal probability distribution for this code, i.e., the values of $P(n), n \in \mathbb{N}_0$.
6. Consider a coding system for non-negative integer numbers, where a certain value n is represented by n "0" bits followed by a "1" bit. For what probability distribution, $P(n), n \in \mathbb{N}_0$, is this coding system maximally efficient?
7. Consider the following table (incomplete) of correspondences between symbols σ_i, probabilities p_i, and codewords:

$$\begin{array}{lcr} \sigma_1(p_1 = 0.30) & \longrightarrow & 00 \\ \sigma_2(p_2 = 0.30) & \longrightarrow & ? \\ \sigma_3(p_3 = 0.15) & \longrightarrow & 110 \\ \sigma_4(p_4 = 0.15) & \longrightarrow & ? \\ \sigma_5(p_5 = 0.10) & \longrightarrow & 111 \end{array}$$

   (a) Complete the table in order to obtain a Huffman code.
   (b) Calculate the redundancy of the code, considering the first order entropy of the information source.
8. Consider the following table of correspondences between symbols σ_i, probabilities p_i, and codewords:

$$\begin{aligned} \sigma_1(p_1 = 0.1) &\longrightarrow 000 \\ \sigma_2(p_2 = 0.2) &\longrightarrow 01 \\ \sigma_3(p_3 = 0.2) &\longrightarrow 10 \\ \sigma_4(p_4 = 0.5) &\longrightarrow 1 \end{aligned}$$

(a) Assuming that the symbols occur independently, calculate the entropy of this information source.

(b) This variable-length code is not built correctly. Why?

(c) Show that it is not possible to build a prefix-free code with this set of codeword lengths.

(d) Propose a variable-length code appropriate for the given probability distribution.

(e) Calculate the redundancy (in relation to the entropy of the source) of the code built in the previous question.

9. Explain, briefly, the principles of one of the coding algorithms based on dictionaries, pointing out the main advantages and disadvantages.

10. Explain, briefly, the working principle of an arithmetic encoder.

11. Give the code sequence produced by a LZ78 encoder, if the input sequence is:

    zzxyxyxxxyxzxxxyxxzzxyxx

    Also, provide the final state of the dictionary.

12. Give the code sequence produced by a LZ77 encoder if the input sequence is:

    zzxyxyxxxyxzxxxyxxzzxyxx

    Consider that the input buffer has size 4 and the dictionary window has size 12.

13. Give the code sequence produced by a LZW encoder if the input sequence is:

    zzxyxyxxxyxzxxxyxxzzxyxx

    Consider that the input alphabet is $\Sigma = \{\text{x, y, z}\}$.

14. Taking into consideration the principles of arithmetic coding, and that $P(0) = P(1) = 0.5$, indicate a value in the $[0, 1)$ interval that represents all messages that start with the binary sequence "1011".

15. Consider that a certain second-order finite-context model (i.e., that uses the two previously occurred symbols for conditioning the probability of occurrence of the next symbol) has already observed the following binary sequence:

    100101101111010010011000

    Indicate an estimate for the probability given by this model that the next symbol will be "1".