MATH - MTHE 474/874 - Information Theory

Fall 2019

Homework # 3

Due Date: Friday November 1, 2019

Material: *Lossless data compression.*

Readings: Section 3.3 of the textbook.

The referred problems are from the textbook.

(1) Player A chooses some object in the universe, and player B attempts to identify the object with a series of yes-no questions. Suppose that player B is clever enough to use the code achieving the minimal expected length with respect to player A's distribution. We observe that player B requires an average of 35.5 questions to determine the object. Find a rough lower bound to the number of objects in the universe.

(2) Consider a DMS with an alphabet of size five. Assume that the source is encoded via a D-ary first-order uniquely decodable code with codeword lengths: $l_1 = l_2 = 1$, $l_3 = l_4 = 2$ and $l_5 = 3$. Find the smallest possible value of D.

(3) Consider a DMS $\{X_i\}$ with alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_8\}$ and symbol probabilities given by $P_X(a_1) = 0.2$, $P_X(a_2) = P_X(a_3) = 0.15$, $P_X(a_4) = P_X(a_5) = P_X(a_6) = P_X(a_7) = P_X(a_8) = 0.1$. Construct two optimal (first-order) ternary prefix codes for the source with different codeword length variances. Compare their average codeword lengths and length variances and explain which code would be preferable in practice.

(4) Problem 3.6.

(5) Problem 3.14.

(6) Consider a stationary Markov source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = \{a, b, c\}$ and transition matrix $Q = [p_{xy}]$ given by

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

where the xy^{th} entry is $p_{xy} = \Pr\{X_2 = y | X_1 = x\}$, $x, y \in \mathcal{X}$.

(a) Design an optimal first-order *ternary* prefix code for this source (i.e., for $n = 1$).

(b) Design an optimal second-order *ternary* prefix code for this source (i.e., for $n = 2$).

(c) Describe and evaluate three appropriate metrics to assess the above two code designs. Which code would you recommend ? Justify your answer.

(7) Consider the stationary Markov source described in Problem 7 of the previous assignment (HW # 2). If $\rho = 1/2$ and $\delta = 3/2$, design first-, second-, and third-order binary Huffman codes for this source. Determine in each case the average lengths per source sample (or average code rates) and compare them to the entropy rate.

Additional Problems for MATH 874 students:

(8) Problem 3.17.

(9) Problem 3.21.

(10) Problem 3.22.
