MATH - MTHE 474/874 - Information Theory

Fall 2021

Homework # 3

Due Date: Friday November 5, 2021

Material: *Lossless data compression.*

Readings: Section 3.3 of the textbook.

The referred problems are from the textbook.

(1) For each D-ary variable-length code $\mathcal{C}$ below, determine (with justification) whether or nor it is uniquely decodable.

(a) $D = 2$, $\mathcal{C} = \{10, 010, 101\}$.

(b) $D = 2$, $\mathcal{C} = \{0, 01, 011, 0111\}$.

(c) $D = 3$, $\mathcal{C} = \{21, 20, 201, 202, 212\}$.

(d) $D = 3$, $\mathcal{C} = \{1, 21, 221, 002, 021, 001\}$.

(e) $D = 4$, $\mathcal{C} = \{10, 12, 13, 22, 121, 133, 220, 221, 223\}$.

(2) Consider the following set of integers: $\{l_1 = 1, l_2 = 1, l_3 = 2, l_4 = 2, l_5 = 3, l_6 = 3, l_7 = 4\}$.

(a) Find the smallest integer $D \geq 2$ such integers $l_1, \ldots, l_7$ satisfy Kraft's inequality in base D.

(b) For the value of D found in part (a), find integer l_7^* so that $l_1, \ldots, l_6, l_7^*$ satisfy Kraft's inequality (in base D) with *equality*.

(c) For the value of D found in part (a), design a D-ary prefix code $\mathcal{C}$ with the integers $l_1, \ldots, l_6, l_7^*$ as its codeword lengths.

(d) Propose a discrete memoryless source for which the above code $\mathcal{C}$ is absolutely optimal (i.e., its average codeword length equals the source's entropy) and compute the source's entropy specifying its units.

(3) Determine (with justification) whether each of the following statements is *True* or *False*:

(a) There does not exist a binary prefix code of size three and with codeword lengths $l_1 = l_2 = 2$ and $l_3 = 3$ for which the following sequence

$$00101100010010110$$

is a concatenation of its codewords (and is hence uniquely decodable).

(b) For a uniformly distributed memoryless source with alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_9\}$, it is not possible to design a ternary first-order Huffman code $\mathcal{C}$ with length variance $\mathrm{Var}(l(\mathcal{C})) = 0$.

(c) Consider a discrete memoryless source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ and probability distribution

$$[p_1, p_2, p_3, p_4, p_5, p_6] = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right],$$

where $p_i := P(X = a_i)$, $i = 1, \cdots, 6$. The binary first-order Shannon code (described in Lecture 16) for the source is optimal.

(4) Consider a discrete memoryless source $\{X_n\}_{n=1}^{\infty}$ with alphabet $\mathcal{X} = \{a, b, c, d, e, f, g\}$. Let p_1 and p_2 be two possible probability mass functions for the source that are described in the table below. Also, let $\mathcal{C} = f(\mathcal{X})$ where $f \colon \mathcal{X} \to \{0, 1, 2, 3\}^*$, be a quaternary (first-order) prefix code for the source as shown below.

Symbol $x \in \mathcal{X}$	$p_1(x)$	$p_2(x)$	$f(x)$
a	1/4	1/4	0
b	1/4	1/8	1
c	1/4	1/8	2
d	1/16	1/8	30
e	1/16	1/8	31
f	1/16	1/8	32
g	1/16	1/8	33

Let $\overline{R}(\mathcal{C}, p_j)$ denote the average compression rate of code $\mathcal{C}$ for the source under distribution p_j, $i, j = 1, 2$.

(a) Calculate $\overline{R}(\mathcal{C}, p_1)$, and $\overline{R}(\mathcal{C}, p_2)$ and compare them to their respective theoretical limit (using the appropriate units).

(c) Which distribution is preferable for the source? Comment qualitatively.

(5) Consider a stationary Markov source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = \{a, b\}$ and transition distribution $P_{X_2|X_1}(a|a) = P_{X_2|X_1}(b|b) = \alpha$, where $1/3 < \alpha < 1/2$.

(a) Design a first-order and a second-order optimal binary variable-length codes for the source and compare their average coding rates.

(b) Determine the limit of the average coding rate of an n-th order optimal binary variable-length code for the source as $n \to \infty$. Justify your answer.

(6) We are interested in constructing first-order binary uniquely decodable (UD) and suffix variable-length codes (recall that a suffix code has the property that none of its codewords is a suffix of another) for a discrete memoryless source (DMS) $\{X_i\}$ with alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_M\}$ and symbol probabilities given by $p_i := P_X(a_i) > 0$, $i = 1, 2, \ldots, M$. The criterion in designing the codes is that they minimize the following expected cost function

$$\bar{L} = \sum_{i=1}^{M} p_i c(l_i),$$

where $l_i, \ldots, l_M$ are the lengths of the codewords for source symbols $a_1, \ldots, a_M$, respectively, and $c : \{1, 2, \ldots\} \to (0, \infty)$ is a (not necessarily linear) cost function. Define

$$\bar{L}_{\text{S}} := \min_{\text{all suffix codes}} \bar{L}$$

and

$$\bar{L}_{\text{UD}} := \min_{\text{all uniquely decodable codes}} \bar{L}.$$

(a) Compare $\bar{L}_{\text{S}}$ and $\bar{L}_{\text{UD}}$.

(b) Describe and evaluate appropriate metrics to assess the performance and complexity of D-ary nth-order UD code designs for a DMS.

(7) Answer the following problems.

(i) A large data set is generated by a stationary source $\{X_n\}_{n=1}^{\infty}$ with finite alphabet $\mathcal{X}$ and joint distribution $P_{X^n}(x^n)$, $x^n \in \mathcal{X}^n$, $n \geq 1$. However the underlying (true) distribution P_{X^n}, being unknown, the data source is approximated by another stationary source $\{\hat{X}_n\}_{n=1}^{\infty}$ with *known* distribution $P_{\hat{X}^n}(x^n)$, $x^n \in \mathcal{X}^n$, $n \geq 1$. (We assume that $P_{X^n}(x^n) > 0$ and $P_{\hat{X}^n}(x^n) > 0$ for all $x^n \in \mathcal{X}^n$.)

For any fixed integer $n \geq 1$, we design an n-th order binary prefix code $f_n : \mathcal{X}^n \to \{0,1\}^*$ for the source under the model $\{\hat{X}_n\}$ by using Shannon's assignment rule to the lengths $l(c_{x^n})$ of codewords $c_{x^n} = f_n(x^n)$, given as follows (cf., Lecture 15):

$$l(c_{x^n}) = \lceil -\log_2 P_{\hat{X}^n}(x^n) \rceil, \qquad x^n \in \mathcal{X}^n.$$

(a) Show that the *true* average compression rate $\bar{R}_n := \frac{1}{n}\sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) l(c_{x^n})$ of the code, computed under the *true* source distribution P_{X^n} satisfies

$$\frac{1}{n}H(X^n) + \frac{1}{n}D(P_{X^n}\|P_{\hat{X}^n}) \leq \bar{R}_n < \frac{1}{n}H(X^n) + \frac{1}{n} + \frac{1}{n}D(P_{X^n}\|P_{\hat{X}^n})$$

and hence estimating the true source distribution P_{X^n} via $P_{\hat{X}^n}$ results in a (inefficiency) cost of $\frac{1}{n}D(P_{X^n}\|P_{\hat{X}^n})$ in the average compression rate.

(b) Now assume that the approximating source $\{\hat{X}_n\}$ is a stationary Markov source $\{\hat{X}_n\}_{n=1}^{\infty}$ with transition distribution $P_{\hat{X}_2|\hat{X}_1}(b|a)$, $a, b \in \mathcal{X}$, obtained by estimating the relative frequencies of transitions (between consecutive data symbols) in the data set. Determine $\lim_{n\to\infty} \bar{R}_n$ in terms of the cross-entropies $H(P_{X_1,X_2}; P_{\hat{X}_1,\hat{X}_2})$ and $H(P_{X_1}; P_{\hat{X}_1})$.

(ii) [**MATH 874 only**] Problem 3.2.