Indian Institute of Technology Bombay

Department of Electrical Engineering

Handout 9 EE 708 Information Theory and Coding

Practice Questions Feb 18, 2022

Notations:

$$U_j^k := U_j, \cdots, U_k,\ j \le k$$
$$U^k := U_1, \cdots, U_k,\ k \ge 1$$
$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Note: If your solution requires any additional assumptions, list those at the start of the answer.

Question 1) Consider two terminals A and B. Terminal A observes the random source process $U_n, n \ge 1$, which are the IID realizations of a binary random variable U with $P(U = 0) = \rho$. The observations at Terminal B are given by $V_n = U_n \oplus Z_{1n}$, where Z_{1n} are IID realizations of Z_1 with $P(Z_1 = 0) = \rho_1$, independent of $U_n, n \ge 1$. The sign $\oplus$ stands for binary addition (XOR).

Assume now the presence of a DMC between A and B, where X is the input to the channel, and $Y = X \oplus Z_2$ is the output. Here, Z_2 is generated IID with $P(Z_2 = 0) = \rho_2$, independent of both (U_n, V_n), as well as the transmitted symbols,

At each time instant, the source produces a new U symbol, and we can transmit once over the channel as well.

(a) Compute $H(U|U \oplus Z_1)$ for the distributions given above, with U independent of Z_1.

[5 marks]

(b) If $\rho_1 = \frac{1}{2}$, what is the maximal value of $\rho \in [0, \frac{1}{2}]$ such that the source U_n can be conveyed to Terminal B with an arbitrary small error probability? (In other words, $P(U^N \ne \hat{U}^N)$ should be made arbitrarily small, possibly by increasing N, where $\hat{U}^N$ is the estimate of U^N at Terminal B)

[5 marks]

(c) Now take $\rho_1 < \frac{1}{2}$, and consider the following communication scheme. Assume that at the start we fix N, and uniformly and independently assign a coupon number $B(u^N) \in \{1, \cdots, M\}$ to each possible u^N sequence. The coupon number of each sequence is revealed to both the terminals.

The scheme proceeds as follows. Suppose Terminal A observes $U^N = u^N$. It attempts to convey the coupon number $B(u^N)$ to Terminal B, using N channel uses of the available DMC. After figuring out the coupon number $B(u^N)$ at the transmitter, Terminal B picks, if available, the unique u^N sequence having a coupon number $B(u^N)$ such that u^N is jointly typical with the corresponding V^N sequence locally available. If no unique u^N jointly typical with V^N exist, an error is declared. Notice that you are free to pick convenient values of N, M etc.

If $\rho_1 < \frac{1}{2}$, what is the maximal value of $\rho \in [0, \frac{1}{2}]$ such that the source U_n can be conveyed to Terminal B with an arbitrarily small error probability? (In the sense defined in part (b)). The answer can be left as an equation in terms of the other given parameters.

[10 marks]

Question 2) Consider a uniform random variable $W \in \{1, \cdots, M\}$, and an IID random sequence $S_n, 1 \le n \le N$, which is independent of W. Here N is a sufficiently large number. For each $1 \le n \le N$, a random variable Y_n is generated as a function of W, S^N and other possible sources of independent randomness. Let $U_n := (W, Y^{n-1}, S_{n+1}^N)$. Relate the LHS and RHS of [?] in the expression given below, and give the conditions for equality.

$$I(W; Y^N) + \sum_{i=1}^{N} I(U_i; S_i) \;\; [?] \;\; \sum_{i=1}^{N} I(U_i; Y_i).$$

Hint: Start with expanding a suitable mutual information term, then bring the remaining random variables to the picture by adding/substracting mutual information terms, that we can apply formulas derived in the course EE708. All undefined variables can be taken as zero.

[10 marks]

Question 3) The redundancy of a source code for the random variable X is defined as the difference between the expected length and entropy, i.e. the redundancy for a source distribution $p(x)$ is

$$R(p) = \sum_{x \in \mathcal{X}} p(x) l(x) - H(X),$$

where $l(x)$ is the length for symbol $x \in \mathcal{X}$ after encoding. Assume the source to have $K = u\,2^k$ symbols, where k is an integer and $1 \le u < 2$.

(a) Give an upper bound to the redundancy of a Huffman code for an equiprobable distribution on the above source. (Tighter bounds will get more marks).

[6 marks]

(b) Note that $H(X)$ is concave in $p(x)$, while the expected length is linear. An engineer claims that the redundancy of the Huffman code designed for any distribution $p(x)$ will become higher if the actual source distribution encountered turns out to be $q(x)$, which is not the same as $p(x)$. Is he/she right or wrong? (justify).

[4 marks]

Question 4) *From Home Work:* The essential ideas of this question are borrowed from Homework 2.

(a) Consider a continuous valued random variable X of finite mean μ, admitting a density $f_X(x)$. Show that the differential entropy $h(X)$ does not depend on the mean μ.

[2 marks]

(b) Of all continuous valued random variables with variance at most P, find the distribution which maximizes the differential entropy, and compute the maximal entropy.

[8 marks]

Question 5) Consider the pair (X^n, Y^n) of $n-$length random vectors. These are not necessarily independent, and are taken from some joint distribution $P(x^n, y^n)$. Take another arbitrary random variable U and define the quantities

$$\mathcal{I}_1 = \sum_{i=1}^{n} I\left(X_i; Y_1^{i-1} | X_{i+1}^n, U\right)$$

$$\mathcal{I}_2 = \sum_{i=1}^{n} I\left(Y_i; X_{i+1}^n | Y_1^{i-1}, U\right)$$

where $X_{i+1}^n = (X_{i+1}, \ldots, X_n)$ and $Y_1^{i-1} = (Y_1, \ldots, Y_{i-1})$. (Recall that $X^n = X_1^n$, and take $X_i = Y_i = 0$ whenever $i \notin \{1, 2, \cdots, n\}$).

(a) What is the logical relation between $\mathcal{I}_1$ and $\mathcal{I}_2$? (Hint: Try with $U = \emptyset$)

(b) Explicitly identify the conditions for equality between $\mathcal{I}_1$ and $\mathcal{I}_2$.

Question 6) Let $N = x2^k$ be the number of source symbols, where $x \in (0, 1]$ and k is some positive integer. Let $p_1, \ldots, p_N$ be the probability assignments on the source symbols. It is given that $p_N = \rho$, and the remaining symbols are all equiprobable.

What do you think is the least value of ρ such that the resulting Huffman code is the same as that of a source having the same probability for each symbol.

Question 7) Consider a source S which takes values in the set $\{0, 1, 2\}$. We have to encode the source sequence $S_n, n \geq 1$. However, S_n here are not generated IID. More specifically, if the i^{th} source realization S_i is j, the next source symbol S_{i+1} is generated according to the distribution $p_j(s)$, for $j \in \{0, 1, 2\}$, independent of everything else. It is given that $p_1(s) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $p_2(s) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and $p_3(s) = (\frac{3}{8}, \frac{3}{8}, \frac{1}{4})$. Assume S_1 to be chosen uniformly. Our task is to convey the above source sequence to an interested receiver.

(a) Suppose the receiver is already aware of the source symbols S_{mK+1}, where $m = 0, 1, 2, \cdots$ and K is some positive integer (imagine a genie giving this information to the receiver). How many bits per source symbol (on average) are now required to convey the entire source to the receiver in a lossless fashion.

[5 marks]

(b) Assume now that the receiver has no prior idea of any of the source symbols, other than the statistical description given before part (a). How many bits per source symbol are required to convey the source sequence to the receiver.

[5 marks]

Question 8) Let $\bar{M} = \sum_{i=1}^{n} p_i \sqrt{l_i}$ be the expected value of the square-root of the codeword lengths l_i associated with an encoding of a random variable X with distribution p.

(a) Let S_{in} be the set of all instantaneous codes for X and S_{du} be the set of all uniquely decodable codes. Which set is bigger (in the sense of the other being a subset.)

[1 mark]

(b) Let $\bar{M}_1 = \min \bar{M}$ over all instantaneous codes; and let $\bar{M}_2 = \min \bar{M}$ over all uniquely decodable codes (for X in Question 1). What inequality relationship exists between $\bar{M}_1$ and $\bar{M}_2$?

[4 marks]

Question 9) A random variable takes values on an alphabet of K letters, and each letter has the same probability. These letters are encoded into binary words using the Huffman procedure so as to minimize the average codeword length. Let j and x be chosen such that

$K = x2^j$, where j is an integer and $1 \leq x < 2$.

(a) Find the number of codewords having lengths less than j?

[3 marks]

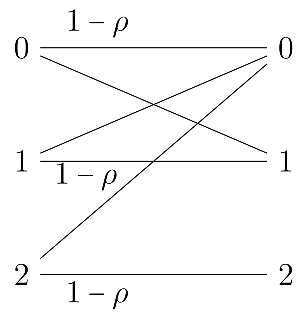
(b) In terms of j and x, how many code words have length j?

[7 marks]

(c) What is the average codeword length?

[5 marks]

Question 10) Consider the following channel



a) Using Fano's inequality, show that the capacity of this channel is at least $C_1(\rho) = \log_2 3 - H_b(\rho) - \rho$.

[5 marks]

b) Is it possible for any channel to have capacity exactly $C_1(\rho)$,

- if YES, find one.
- if NO, explain why.

[5 marks]

Question 11) Let $X_1, X_2, \ldots, X_n$ be (possibly dependent) binary random variables. Suppose one calculates the run lengths $R = (R1, R2, \ldots)$ of this sequence (in the order they occur). For example, the sequence $x = 0001100100$ yields the run lengths $R = (3, 2, 2, 1, 2)$. Compare $H(X_1, X_2, \ldots, X_n)$, $H(R)$ and $H(X_n, R)$. Show all equalities and inequalities and bound all the differences.

[10 marks]

Question 12) Consider a source S with M symbols. We call a distribution $D-$**adic** if every source-symbol probability can be expressed as,

$$P(S_i) = D^{-L_i}, \; 1 \leq i \leq M,$$

where L_i, D are positive integers. As usual, D stands for the cardinality of the output symbols(letters) which form the codebook. Also let M be an integer such that Kraft inequality is always satisfied with equality for $D-$ary Huffman codebooks for S.

a) Let Ω_1 be the class of all $D-$adic distributions on the source S. Show that for every distribution $p \in \Omega_1$, symbol-wise Huffman coding is at least as good as any asymptotic scheme operating on source sequences S^n.

[5 marks]

b) Let Ω be another class of distributions, where each symbol probability is an integer multiple of $D^{-L_{max}}$. Here L_{max} is the maximal code-word length, where the maximum is taken over all p and their corresponding Huffman codebooks. Consider an arbitrary distribution q on the source symbols. Will Huffman coding of q yield the same codebook as that of at least one distribution in Ω? Explain your answer (in not more than a page).

[5 marks]

c) Suppose we have pre-computed the Huffman codebooks for each $p \in \Omega_1$. We have now stored this codebooks in some compact memory onboard in a sensing device. Let the source S be distributed as q, where q may not be in Ω_1. We don't have the means (inside the sensor) to compute the Huffman code for q, rather we will choose one from the codebooks we made for Ω_1.

Show that the best choice(for symbol-wise source coding) is to pick the codebook for q^*, where

$$q^* = \underset{p \in \Omega_1}{\operatorname{argmin}} D(q||p)$$

[10 marks]

d) Let the source distribution q be unknown at the encoder, but it is known that there exist Huffman codes which can compress this source to $\log_D M - c$ output letters per symbol on the average, where c is some positive constant (not too small, say ≈ 0.25). Will you prefer to use the codebook with all M equi-length codewords for this source. Explain your answer in less than 3 lines (marks are for the explanation).

[5 marks]