Information and Communication (EC4.102)
Assignment 3
Spring-24, March 25, 2024

---

1. Complete the assignment all by yourself. You can discuss approaches to the questions but the solution should be yours.
2. Clearly state the assumptions made (if any) that are not specified in the question.

Deadline: 5th April 8:00 PM　　　　Max. Marks : 71

---

1. Given that $F(x)$ is a valid Cumulative Distribution Function, which of the following is a valid Cumulative Distribution function? $[3 \times 2 = 6]$
   (a) $\alpha F(x)$ where $\alpha \in \mathbb{R}$
   (b) $F(x)^2$
   (c) $F(x) + (1 - F(x)) \log(1 - F(x))$. [Hint: What properties must a function satisfy to be called as a CDF? Check all these.]

2. For a continuous random variable X, and a non-zero constant $a \in \mathbb{R}$, derive the formulae for the following: $[4 \times 2 + 2 = 10]$
   (a) $\mathbb{E}[X + a]$
   (b) $\mathbb{E}[aX]$
   (c) $var(X + a)$
   (d) $var(aX)$

   Recall that the expectation of a random variable is essentially a weighted average. The variance in the random variable measures the average squared-deviation from the mean (i.e., it measures how much the values of the random variable deviates from the mean, on average). Explain the intuitive understanding you have, for each of the above expressions you have calculated.

3. A seven-match cricket series is played between two teams, **A** and **B**. A team wins the series as soon as they win 4 matches. Let X be the random variable that represents the outcome of the series after completion (i.e., either team has won). (For example, X can be ABBAAA, ABABBAA, BBABB etc.) Let the random variable Y be the number of games played which ranges from 4 to 7.
   (a) Understand that the quantities $H(Y|X)$ is measuring the uncertainty in Y conditioned on the fact that X is known. Without doing any numerical calculations, what do you intuitively think $H(Y|X)$ is? Explain with reasons. [2]
   (b) Understand that the the quantities $H(X)$ and $H(Y)$ measure the uncertainity in X and Y. Which of them do you think is larger? Why? [2]
   (c) Assuming that the probability of winning is the same for both teams and the outcome of each match is independent of the others, calculate $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$ and $I(X;Y)$. [Hint: Find out in how many ways can a series which ends in 4,5,6,7 matches be played. This will give you the probability distribution of X and thereby determining the distribution of Y.] $[5 \times 2 = 10]$

4. For three joint random variables X, Y, Z , show that:

$$2H(X, Y, Z) \leq H(X, Y) + H(Y, Z) + H(X, Z)$$

(Hint: Use the chain rule for entropy on the left and the relation between joint entropy and conditional entropy on the right. Furthermore use the property "Conditioning does not increase entropy".) [3]

5. Consider 2 independent fair binary random variables X and Y (i.e., they take values in $\{0, 1\}$ with equal probabilities.). Let another binary random variable, Z, be defined as $Z = X + Y$. What can you say about $I(X; Y)$ and $I(X; Y|Z)$. Identify which quantity is larger and under what conditions are they equal? Now, suppose that $X = Y$ (i.e., they have the same probability distribution but are not independent), what can you then say about $I(X; Y)$ and $I(X; Y|Z)$. Does the previous relationship still apply? [4]

6. Prove the following inequalities and state the conditions for equality: $[4 + 2 + 3]$
   (a) $H(X_1, X_2, X_3, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$
   (b) $I(X; Y) \geq 0$
   (c) $H(X) \leq log|\mathcal{X}|$

   (Hint: To prove part (b) and (c), use the theorem $D(p||q) \geq 0$. You don't need to prove this. Just use the result as is.)

7. Reduce the following expressions, using the relevant chain rules, to the maximum extent possible. $[2 \times 2 = 4]$
   (a) $H(X, Y, Z|W_1, W_2, W_3)$.
   (b) $I(X_1, X_2; Z_1, Z_2|Y_1, Y_2)$.

8. Assume a source with 10 symbols – $\{A, B, C, D, E, F, G, H, I, J\}$ where, $[4 + 2 + 3]$

$\mathbf{S}$	A	B	C	D	E	F	G	H	I	J
$\mathbb{P}(X = \mathbf{S})$	0.3	0.25	0.2	0.1	0.05	0.03	0.03	0.02	0.01	0.01

   (a) Perform ternary Huffman coding on the given source symbols. Present the source code in a neat tabulated manner.
   (b) Find the expected length $L(C)$ of the source code generated above.
   (c) Find the coding efficiency η.

9. Assume a source X with 4 symbols – $\{x_i : i \in \{1, 2, 3, 4\}\}$, with distribution p_X taking values $0.4, 0.3, 0.2, 0.1$, respectively. $[6 \times 2 = 12]$
   (a) Find the entropy of this source.
   (b) Construct binary Huffman codes for p_X. Calculate its expected length and call it L_1.
   (c) Now, imagine that two i.i.d. (independent and identically distributed) symbols coming out from the same source, denoted by X_1, X_2. What is the joint distribution of these two symbols, which we denote by p_{X_1,X_2}?
   (d) Can you construct an length-optimal source code (i.e., Huffman code) for this distribution in part 3? Calculate its expected length and call it L_2.

(e) Compare the three quantities: the two ratios obtained as the average length of the source code to the number of source symbols (i.e., $L_1/1 = L_1$ and $L_2/2$), along with the entropy $H(X)$. What do you understand from this comparison?

(f) Suppose we compressed 3-iid-symbols using the Huffman code, and got the normalized expected length as L_3? How would you think this would compare with $H(X)$ and L_1 and L_2?
