Indian Institute of Technology Kharagpur
Information and Coding Theory (MA41024/MA60020)
Assignment - 2, Spring Semester 2020-21

Last Date of submission: March 8, 2021

1. Consider the following code with source alphabet $S = \{a, b, c\}$ and code alphabet $\mathcal{D} = \{0, 1\}$.

$$a = 0,\ b = 01,\ c = 011.$$

   Then decide whether the code is uniquely decodable and prefix.

2. A code is called lossless if distinct source letters have distinct codewords.
   (a) Then produce an example of a lossless code with code alphabet $\mathcal{D} = \{0, 1, \ldots, 2\}$ such that Kraft inequality is violated.
   (b) Give an example of a lossless code with variable length codewords such that the expected codeword length is strictly less than the entropy of the information source.

3. True or False (Justify your answer): Given a prefix code with code alphabet $\{0, \ldots, K\}$ it is always possible to define another prefix code such that $p(j) \geq p(i)$ implies $|c(j)| \leq |c(i)|$ without any increment of the expected codeword length of the original code, where $|c(x)|$ denotes the codeword length of the codeword of x and $p(x)$ is the probability of the source letter x.

4. Shannon-Fano coding scheme: Let $\mathcal{D} = \{1, 2, \ldots, K\}$. Then fix the codeword length n_k of $k \in \mathcal{D}$ such that

$$D^{-n_k} \leq p(k) < D^{-n_k+1}, 1 \leq k \leq K.$$

   Then a prefix code with shortest n_k upwards, is known as the Shannon-Fano code. Then answer the following:
   (a) Which result/theorem guarantees the existence of such a prefix code.
   (b) Design a Shannon-Fano code for the source alphabet $\mathcal{D} = \{a, b, c, d, e\}$ with $p(a) = 0.45$, $p(b) = 0.25$, $p(c) = 0.2$, $p(d) = 0.05$, and $p(e) = 0.05$
   (c) Is the Shannon-Fano code optimal? Justify your answer.

5. Let X and Y denote the input and output random variables for a channel. Define the error probability that two outcomes x and y are different as

$$P_e = \sum_x \sum_{y \neq x} P(x, y).$$

   Then consider the joint random variable $\mathbf{X}^L = (X_1, X_2, \ldots, X_L)$ where X_l is a copy of X, $1 \leq l \leq L$. Then show that

$$\frac{1}{L} \sum_{l=1}^{L} \mathcal{H}(P_{e,l}) \leq \mathcal{H}(\langle P_e \rangle),$$

   where $\langle P_e \rangle = \sum_{l=1}^{L} P_{e,l}/L$ and $P_{e,l}$ denotes the error probability for $x_l \neq y_l$ for the joint random variable $(\mathbf{X}^L, \mathbf{Y}^L)$.

6. Let a_1, a_2 denote the source letters for a channel, and the output letters are $a_1, a_2, -$. Suppose the channel is defined by the transition probabilities $p(a_1|a_1) = p(a_2|a_2) = 1/2$, $p(-|a_1) = p(-|a_2) = 1/2$. Then derive the channel matrix and the capacity of the channel.

7. Consider the random variable

$$X = \begin{matrix} x = & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ p(x) = & 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{matrix}.$$

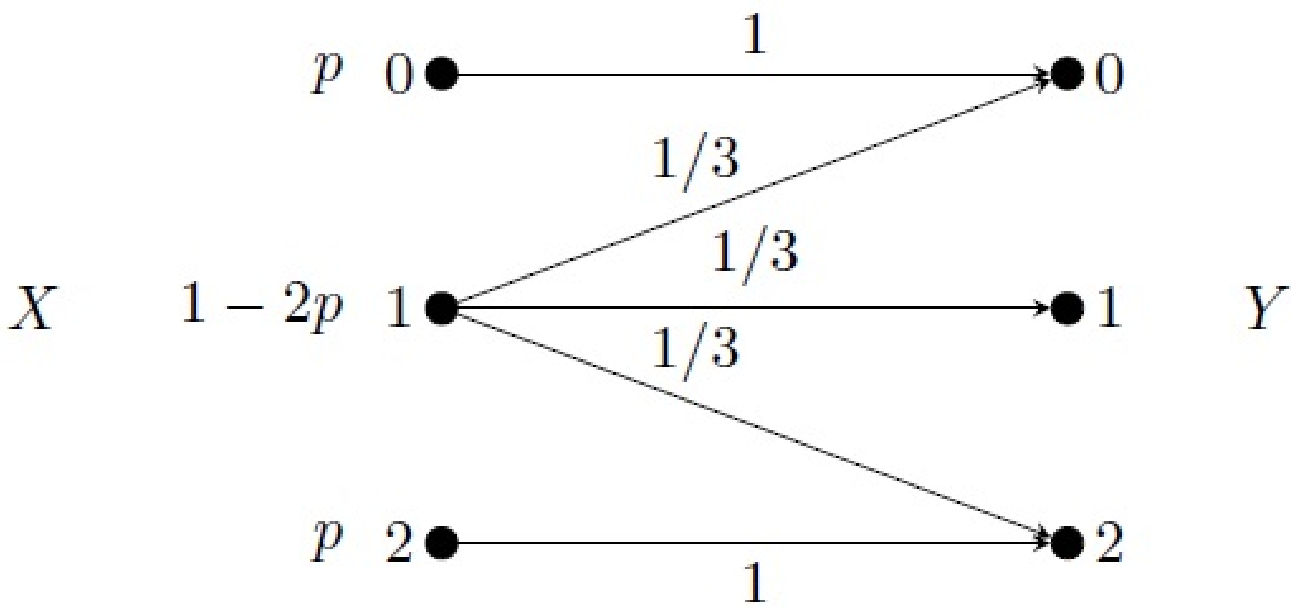

Figure 1: Channel

(a) Find a binary Huffman code for X.

(b) Find the expected codelength for this encoding.

(c) Find a ternary Huffman code for X.

8. State source coding and channel coding theorems, and justify the theorems with examples.

9. Calculate the channel capacity of the channel given in Figure 1.

10. State and give an interpretation of Fano's lemma.

All the best!!