Indian Institute of Technology Kharagpur
Information and Coding Theory (MA41024/MA60020)
Assignment - 1, Spring Semester 2021-22

Last date of submission ($\star$-marked questions): February 6, 2022

1. The input source of a noisy communication channel is a random variable X with alphabet $\{1, 2, 3, 4\}$. The output random variable for the channel is denoted as Y. Suppose the joint pmf of the pair (X, Y) is given by

Y \ X	1	2	3	4
1	1/8	0	0	0
2	0	1/8	1/4	1/8
3	0	1/8	1/8	0
4	0	1/8	0	0

   Then (Everything is to be calculated in bits)

   (a) Determine $H(X)$, $H(Y)$, $H(X,Y)$, $H(Y|X)$, and $H(X|Y)$.

   (b) Calculate $I(X;Y)$. Justify how $I(X;Y)$ and $I(x;y)$, where all pair of symbols (x, y).

2. $\star$ The probability of being male is 0.5 and for being female is also 0.5. Suppose that 20% of males and 6% of females are taller than 6 feet. Calculate the probability that if somebody is taller than 6 ft, that person must be male. Calculate the amount of information you gain (in bits) by learning that a man is taller than 6 ft. Calculate the information that you gain by learning that a female is tall. Finally, how much information do you gain from learning that a tall person is female?

3. Calculate the entropy in bits for each of the following random variables:

   (a) Pixel values in an image whose possible grey values are all the integers from 0 to 255 with uniform probability.

   (b) Humans grouped by whether they are, or are not, mammals.

   (c) Gender in a tricolored insect population whose three color occur with probabilities $1/4, 1/4$, and $1/2$.

(d) A population of persons classified by whether they are older, or not older, than the population's median age.

4. Consider a binary symmetric communication channel, whose input source is the alphabet $X = 0, 1$ with probabilities $0.3, 0.7$ whose output alphabet is Y $= 0, 1$ with transmission error probability ϵ.

   (a) What is the entropy of the source $H(X)$?

   (b) What is the probability distribution of the outputs , and the entropy of this output distribution $H(Y)$?

   (c) What is the joint probability distribution for the source and the output, and what is the joint entropy $H(X, Y)$?

   (d) What is the mutual information of this channel, $I(X; Y)$?

   (e) How many values are there for ϵ for which the mutual information of this channel is maximal? What are those values, and what then is the capacity of such a channel in bits? What is the physical interpretation of the result obtained?

5. $\star$ Suppose that crows are black with probability 0.6, that they are male with probability 0.5 and female with probability 0.5, but that male crows are 3 times more likely to be black than are female crows. If you see a non-black crows, what is the probability that it is male? How many bits worth of information are contained in a report that a non-black crows is male?

   Rank-order for this problem, from greatest to least, the following uncertainties: (i) uncertainty about colour; (ii) uncertainty about gender; (iii) uncertainty about colour, given only that a crow is male; (iv) uncertainty about gender, given only that a crow is non-black.

6. Consider n different discrete random variables, named $X_1, X_2, ..., X_n$, each of which has entropy $H(X_i)$, $i = 1, \dots, n$. Suppose that random variable X_j has the smallest entropy, and that random variable X_k has the largest entropy.

   (a) What is the upper bound on the joint entropy $H(X_1, X_2, ..., X_n)$ of all these random variables? Under what condition will this upper bound be reached?

   (b) What is the lower bound on the joint entropy $H(X_1, X_2, ..., X_n)$ of all these random variables? Under what condition will the lower bound be reached?

7. Suppose that X is a random variable whose entropy $H(X)$ is 3 bits. Suppose that $Y(X)$ is a deterministic function that takes on a different value for each value of X.

   (a) What then is $H(Y)$, the entropy of Y ?

   (b) What is $H(Y|X)$?

   (c) What is $H(X|Y)$?

   (d) What is $H(X, Y)$?

(e) Suppose now that the deterministic function $Y(X)$ is not invertible; i.e., different values of X may correspond to the same value of $Y(X)$. In that case, what could you say about $H(Y)$?In that case, what could you say about $H(X|Y)$?

8. $\star$ Let P, Q and R be probability distributions on $S = \{s_1, s_2, \ldots, s_n\}$, with $P(s_i) = p_i$, $Q(s_i) = q_i$ and $R(s_i) = r_i$, $1 \leq i \leq n$. Then show that $D(P||R) = D(P||Q) + D(Q||R)$ if and oly if

$$\sum_{i=1}^{n} (p_i - q_i) \log(q_i/r_i) = 0.$$

9. Prove that the function $\rho(X,Y) = H(X|Y) + H(Y|X)$ is a metric on the set of all random variables defined on same sample space. Prove that

$$\begin{aligned}\rho(X,Y) &= H(X) + H(Y) - 2I(X;Y)\\ &= H(X,Y) - I(X;Y)\\ &= 2H(X,Y) - H(X) - H(Y).\end{aligned}$$

10. $\star$ Write a python code to plot the entropy surface for the random variable X associated with the sample space $\mathcal{X} = \{a, b, c\}$ with pmf $p(a) = k$, $p(b) = l$, $p(c) = 1 - k - l$ for different values of k, l. Can you locate the maximum entropy from the plotted surface? Justify your answer.

11. $\star$ A channel has binary input and output alphabets and transition probabilities given by the following matrix: $Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$. Find the maximum value which the mutual information can attain and the maximizing input probability distribution.

12. $\star$ Suppose X, Y and Z are discrete random variables. Then show the following:

(a) $I(X;Y|Z) = I(Y;X|Z)$

(b) $I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$

(c) $I(X;Y|Z) \geq 0$ if and only if (X,Z,Y) is a Markov chain.

13. $\star$ The set ${A_\epsilon}^{(n)}(X,Y)$ of jointly typical sequences $\{(x^n, y^n)\}$ with respect to the distribution $p(x,y)$ is the set of n-sequences with empirical entropies ϵ-close to the true entropies:

${A_\epsilon}^{(n)}(X,Y) = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : |-\frac{1}{n}log(x^n) - H(X)| < \epsilon, |-\frac{1}{n}log(y^n) - H(Y)| < \epsilon, |-\frac{1}{n}log(x^n, y^n) - H(X,Y)| < \epsilon\}$ where $p(x^n, y^n) = \Pi_{i=1}^{n} p(x_i, y_i)$ for preassigned $\epsilon > 0$.

We call the sequence x^n typical for ${A_\epsilon}^{(n)}(X) = |-\frac{1}{n}log(x^n) - H(X)| < \epsilon$ for a preassigned $\epsilon > 0$. Similarly it is defined for y.

Now, Consider a binary symmetric channel with transition error probability 0.1. The input distribution is a uniform distribution $[i.e., p(x) = (1/2, 1/2)]$, which yields the joint distribution p(x, y). Then

(a) Calculate $H(X), H(Y), H(X,Y)$, and $I(X,Y)$ for the joint distribution above.

(b) Let $X_1, X_2, ..., X_n$ be drawn i.i.d. according the Bernoulli(1/2) distribution. Of the 2^n possible input sequences of length n, which of them are typical [i.e., member of ${A_\epsilon}^{(n)}(X)$ for $\epsilon = 0.2$]? Which are the typical sequences in ${A_\epsilon}^{(n)}(Y)$?

All the Best!!