EE276: Homework #4

Due on Friday Feb 13, 6pm - Gradescope entry code: E6VP4X

1. **Maximum Differential Entropy**

   (a) Show that among all distributions supported in an interval [a,b], the uniform distribution maximizes differential entropy.

   (b) Let X be a continuous random variable with $\mathcal{E}[X^4] \leq \sigma^4$ and Y be a continuous random variable with a probability density function $g(y) = c \exp\left(-\frac{y^4}{4\sigma^4}\right)$ where $c = \frac{1}{\int_{-\infty}^{\infty} \exp\left(-\frac{y^4}{4\sigma^4}\right) dy}$. Show that

   $$h(X) \leq h(Y)$$

   with equality if and only if X is distributed as Y.
   [Hint: you can use the fact that $\mathcal{E}[Y^4] = \sigma^4$.]

2. **Channel capacity.**
   Find the capacity of the following channels with given probability transition matrices, where the element p_{ij} of the matrix represents $p(y = j|x = i)$:

   (a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

   (b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

   $$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

   (c) $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ (the Z-channel)

   $$p(y|x) = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

3. **Choice of channels.**
   Let $\mathcal{C}_1 \equiv (\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $\mathcal{C}_2 \equiv (\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$ be two channels with capacities C_1, C_2 respectively. Assume the input and output alphabets for the two channels are disjoint, i.e., $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$. Consider a channel $\mathcal{C}$, which is a union of the two channels $\mathcal{C}_1$, $\mathcal{C}_2$, where at each time, one can send a symbol over $\mathcal{C}_1$ or over $\mathcal{C}_2$ but not both. In this problem, we calculate the capacity of $\mathcal{C}$ in terms of the C_1 and C_2.

Define θ as an indicator random variable denoting whether channel $\mathcal{C}_1$ or $\mathcal{C}_2$ is used in a particular transmission. Let X and Y denote the input and output for the channel $\mathcal{C}$. Note that X follows a mixture distribution over the (disjoint) alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$:

$$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1-\alpha \end{cases}$$

where X_1 and X_2 are random variables taking values in $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively.

(a) Argue that $H(\theta|X) = H(\theta|Y) = 0$.

(b) Show that

$$I(X;Y) = h_2(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha)I(X_2;Y_2)$$

where Y_1 and Y_2 are the channel outputs when X_1 and X_2 are transmitted through $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.
Hint: Start with $I(Y;X,\theta) = I(Y;\theta) + I(Y;X|\theta) = I(Y;X) + I(Y;\theta|X)$.

(c) Let C be the capacity of the channel $\mathcal{C}$. Maximize the expression in (b) over α, P_{X_1}, P_{X_2} to show that $2^C = 2^{C_1} + 2^{C_2}$.

(d) Let $C_1 = C_2$. Then show that $C = C_1 + 1$ and give an intuitive explanation.

4. **Cascading an Erasure channel**

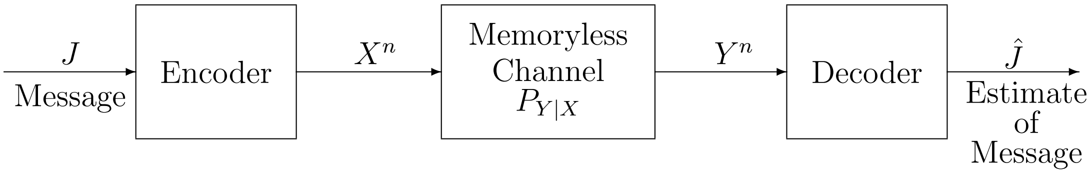
Let $\{\mathcal{X}, p(y|x), \mathcal{Y}\}$ be a discrete memoryless channel with capacity C. Suppose this channel is immediately cascaded with an erasure channel $\{\mathcal{Y}, p(s|y), \mathcal{S}\}$ that erases α of its symbols.

Specifically, $\mathcal{S} = \{y_1, y_2, \ldots, y_m, e\}$, and

$$\begin{aligned} \Pr\{S = y|X = x\} &= (1-\alpha)p(y|x), \quad y \in \mathcal{Y}, \\ \Pr\{S = e|X = x\} &= \alpha. \end{aligned}$$

Determine the capacity of this channel.

5. **Minimizing Channel Probability of Error**.
Below, we are given a communication setting as seen in lecture.



J is a message uniformly distributed on $\{1, 2, ..., M\}$ passed into the system. The encoder maps message J onto its corresponding n-length codeword X^n from codebook $c_n = \{X^n(1), X^n(2), ..., X^n(M)\}$. The encoded message is sent through a memoryless channel characterized by $P_{Y|X}$, and we receive Y^n as output.

The decoder is responsible for estimating J from Y^n; it is a function $\hat{J}$ that maps Y^n to one of the symbols in $\{1, 2, ..., M, \text{error}\}$. We define the probability of error $P_e = P(\hat{J}(Y^n) \neq J)$. Show that P_e, for a fixed codebook c_n, is minimized by:

$$\hat{J}(y^n) = \text{argmax}_{1 \leq j \leq M} P(J = j | Y^n = y^n).$$

6. **Output power constraint**. Consider an additive white Gaussian noise channel with an expected *output* power constraint P. Thus $Y = X + Z$, $Z \sim N(0, \sigma^2)$, Z is independent of X, and $EY^2 \leq P$. Find the channel capacity.