EE276: Homework #2

Due on Friday Jan 23, 6pm - Gradescope entry code: E6VP4X

1. **Data Processing Inequality.**
The random variables X, Y and Z, belonging to alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ respectively, form a Markov triplet $(X - Y - Z)$ if $p(z|y) = p(z|y, x)$, or, equivalently, if $p(x|y) = p(x|y, z)$. If X, Y, Z form a Markov triplet $(X - Y - Z)$, show that:

   (a) $H(X|Y) = H(X|Y, Z)$ and $H(Z|Y) = H(Z|X, Y)$

   (b) $H(X|Y) \leq H(X|Z)$

   (c) $I(X;Y) \geq I(X;Z)$ and $I(Y;Z) \geq I(X;Z)$

   (d) $I(X;Z) \leq \log |\mathcal{Y}|$

   (e) $I(X;Z|Y) = 0$

   where the *conditional mutual information* of random variables X and Y given Z is defined by

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \end{aligned}$$

2. **Conditional mutual information vs. unconditional mutual information.** Give examples of joint random variables X, Y and Z such that

   (a) $I(X;Y \mid Z) < I(X;Y)$,

   (b) $I(X;Y \mid Z) > I(X;Y)$.

3. **Prefix and Uniquely Decodable codes**
Consider the following code:

u	Codeword
a	1 0
b	0 0
c	1 1
d	1 1 0

   (a) Is this a Prefix code?

   (b) Argue that this code is uniquely decodable, by describing an algorithm for the decoding.

4. **Relative entropy and the cost of miscoding.** Let the random variable X be defined on $\{1, 2, 3, 4, 5, 6\}$ according to pmf p. Let p and another pmf q be

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	1/2	1/2	0	0
2	1/8	1/4	100	10
3	1/8	1/16	101	1100
4	1/8	1/16	110	1101
5	1/16	1/16	1110	1110
6	1/16	1/16	1111	1111

   (a) Calculate $H(X)$, $D(p||q)$ and $D(q||p)$.

   (b) The last two columns above represent codes for the random variable. Verify that codes C_1 and C_2 are optimal under the respective distributions p and q.

   (c) Now assume that we use C_2 to code X. What is the average length of the codewords? By how much does it exceed the entropy $H(X)$, i.e., what is the redundancy of the code?

   (d) What is the redundancy if we use code C_1 for a random variable Y with pmf q?

5. **Shannon code.** Consider the following method for generating a code for a random variable X which takes on m values $\{1, 2, \ldots, m\}$ with pmf p having probabilities $p_1, p_2, \ldots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define

$$F_i = \sum_{k=1}^{i-1} p_k,$$

   i.e. the sum of the probabilities of all symbols less than i. Then the codeword for i is the number $F_i \in [0, 1]$ rounded off to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$.

   (a) Show that the code constructed by this process is prefix-free and the average length satisfies

$$H(X) \leq L < H(X) + 1.$$

   (b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$.

   (c) Now, suppose the code in (a) is used on a random variable $\tilde{X}$ taking values in $\{1, 2, \ldots, m\}$ distributed with pmf q having probabilities $q_1, q_2, \ldots q_m$. Show that the average length satisfies

$$H(\tilde{X}) + D(q||p) \leq L < H(\tilde{X}) + D(q||p) + 1$$

6. **AEP.** Let X_i for $i \in \{1, \ldots, n\}$ be an i.i.d. sequence from the p.m.f. $p(x)$ with alphabet $\mathcal{X} = \{1, 2, \ldots, m\}$. Denote the expectation and entropy of X by $\mu := \mathbb{E}[X]$ and $H := -\sum p(x) \log p(x)$ respectively.

   For $\epsilon > 0$, recall the definition of the typical set

$$A_\epsilon^{(n)} = \left\{ x^n \in \mathcal{X}^n : \left| -\frac{1}{n} \log p(x^n) - H \right| \leq \epsilon \right\}$$

and define the following set

$$B_\epsilon^{(n)} = \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n} \sum_{i=1}^{n} x_i - \mu \right| \leq \epsilon \right\}.$$

In what follows, $\epsilon > 0$ is fixed.

(a) Does $\mathbb{P}\left(X^n \in A_\epsilon^{(n)}\right) \to 1$ as $n \to \infty$?

(b) Does $\mathbb{P}\left(X^n \in A_\epsilon^{(n)} \cap B_\epsilon^{(n)}\right) \to 1$ as $n \to \infty$?

(c) Show that for all n,

$$|A_\epsilon^{(n)} \cap B_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}.$$

(d) Show that for all n sufficiently large:

$$|A_\epsilon^{(n)} \cap B_\epsilon^{(n)}| \geq \left(\frac{1}{2}\right) 2^{n(H-\epsilon)}.$$

7. **An AEP-like limit and the AEP (Bonus)**

(a) Let $X_1, X_2, \ldots$ be i.i.d. drawn according to probability mass function $p(x)$. Find the limit in probability as $n \to \infty$ of

$$p(X_1, X_2, \ldots, X_n)^{\frac{1}{n}}.$$

(b) Let $X_1, X_2, \ldots$ be an i.i.d. sequence of discrete random variables with entropy $H(X)$. Let

$$C_n(t) = \{x^n \in \mathcal{X}^n : p(x^n) \geq 2^{-nt}\}$$

denote the subset of n-length sequences with probabilities $\geq 2^{-nt}$.

i. Show that $|C_n(t)| \leq 2^{nt}$.

ii. What is $\lim_{n\to\infty} P(X^n \in C_n(t))$ when $t < H(X)$? And when $t > H(X)$?