2 Relative Entropy and Mutual Information [17 Points]

We define **relative entropy** between two discrete distributions $\mathbf{p} \in \{p_1, \cdots, p_n\}$ and $\mathbf{q} = \{q_1, \cdots, q_n\}$ as:

$$D(\mathbf{p}\|\mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

The above quantity is equal to 0 if and only if $\mathbf{p} = \mathbf{q}$. Otherwise, this quantity is positive. We can think of **relative entropy** as a measure of "distance" between two distributions. This quantity is widely known as the **Kullback-Leibler (KL) divergence** between two distributions (you can assume that none of the q_i and p_i are 0).

1. **[5 Points]** Prove that the **KL divergence** is always non-negative. Show all steps of your work
   Hint: You may use this inequality without proof: $x - 1 \geq \log(x)$, where $x \in \mathbb{R}$ with equality if and only if $x = 1$.

$$D(p\|q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

$$\xrightarrow{(-1)} -D(p\|q) = -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

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

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

$$= \sum_{i=1}^{n} p_i \log \underbrace{\frac{q_i}{p_i}}_{x} \leq \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1\right)$$

$$= \sum_{i=1}^{n} \cancel{p_i} \times \frac{q_i}{\cancel{p_i}} - \sum_{i=1}^{n} p_i$$

$$= \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i = 1 - 1 = 0$$

$$\Rightarrow -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \leq 0 \Rightarrow \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \geq 0 \quad \#$$

2. **[5 Points]** True or False: **KL-divergence** is symmetric (i.e. $D(\mathbf{p}\|\mathbf{q}) = D(\mathbf{q}\|\mathbf{p})$). If you think the statement is true, please write down a sketch of proof. Otherwise, give a counterexample. Show all steps of your work

False, counter example :

$p \in \{0.2, 0.8\}$

$q \in \{0.6, 0.4\}$

$$D(p\|q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} = 0.2 \log \frac{0.2}{0.6} + 0.8 \log \frac{0.8}{0.4} \cong 0.1454$$

$$D(q\|p) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} = 0.6 \log \frac{0.6}{0.2} + 0.4 \log \frac{0.4}{0.8} \cong 0.1659$$

$\rightarrow D(p\|q) \neq D(q\|p)$

#

3. **[7 Points]** Let X and Y be random variables taking the values $\{1, \cdots, n\}$. Recall the definition of mutual information between X and Y:

$$I(X;Y) = H(Y) - H(Y|X)$$

Using the conclusion of part 1), show that the mutual information is always non-negative. Show all steps of your work.

$$H(Y) = -\sum_{y \in Y} P(y) \log P(y)$$

$$= -\sum_{y \in Y} \sum_{x \in X} P(x, y) \log P(y)$$

$$H(Y|X) = \sum_{i=1}^{n} P(X=i) H(Y|X=i)$$

$$= \sum_{i=1}^{n} P(X=i) \times \left(-\sum_{j=1}^{n} P(Y=j|X=i) \log_2 P(Y=j|X=i)\right)$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i) P(Y=j|X=i) \log_2 P(Y=j|X=i)$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 P(Y=j|X=i)$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 P(Y=j) + \sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 P(Y=j|X=i)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \left[\log_2 P(Y=j|X=i) - \log_2 P(Y=j)\right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 \frac{P(Y=j|X=i)}{P(Y=j)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 \frac{P(Y=j \mid X=i) P(X=i)}{P(Y=j) P(X=i)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} P(X=i, Y=j) \log_2 \frac{P(Y=j, X=i)}{P(Y=j) P(X=i)}$$

$$= D\left(p(X=i, Y=j) \,\|\, p(X=i) P(Y=j) \right)$$

Based on the conclusion of part 1) : $D(p \| q) \geq 0$

$I(X;Y) = D\left(p(X=i, Y=j) \,\|\, P(X=i) P(Y=j) \right) \geq 0$ ※