Math Review Homework.
Information Theory Problem Set 1.

In the following exercises, $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ are probability densities.

1. Show that the distribution $\mathbf{p}_{max}$ that has the maximum entropy over all distributions $\mathbf{p}$ is the uniform distribution.

2. Calculate the asymetric and symmetric KL-divergences and the Euclidean distances between $\mathbf{p} = 0.2, 0.5, 0.3$ and $\mathbf{q}_1 = (0.1, 0.6, 0.3)$ and between $\mathbf{p}$ and $\mathbf{q}_2 = (0.6, 0.1, 0.3)$.

3. Calculate the values of mutual information between the distributions in the previous problem if the joint distributions are given by

$$f(\mathbf{p}, \mathbf{q}_1) = \begin{bmatrix} 0.06 & 0.08 & 0.06 \\ 0.01 & 0.34 & 0.15 \\ 0.03 & 0.18 & 0.09 \end{bmatrix} \text{ and } f(\mathbf{p}, \mathbf{q}_2) = \begin{bmatrix} 0.12 & 0.02 & 0.06 \\ 0.30 & 0.05 & 0.15 \\ 0.18 & 0.03 & 0.09 \end{bmatrix}$$

4. A function f is called *convex* if $\forall a \in [0,1],\ f(ap_1 + (1-a)p_2) \leq af(p_1) + (1-a)f(p_2)$. A function f is concave if $-f$ is convex.

   (a) Draw a sketch of a convex function and use it to illustrate the definition.

   (b) Is entropy convex or concave or neither? Why or why not?

   (c) Rewrite mutual information using only marginal and conditional probabilities using the notation $p_n = p(x_n)$, $q_n = p(y_n)$ etc.. Is mutual information convex or concave or neither as a function of $p(x_n)$ or as a function of $p(x)$ or of $p(y|x)$.

5. The *conditional entropy* of a random variable *Y* given a random variable *X* is defined to be $H(Y|X) = \sum_x p(x) H(Y|X = x)$.

   (a) Calculate the conditional entropy if $X \sim \mathbf{p}$ and $Y \sim \mathbf{q}_1$.

   (b) Is conditional entropy always smaller or larger than joint entropy (the entropy of the joint distribution)? Why or why not? Your argument for why or why not should be based on mathematics.

6. Construct examples of two pdfs that have zero mutual information and pdfs that do not have zero mutual information. Don't use the pdfs from previous problems.

7. What is the relationship between mutual information and KL-divergence?