Introduction to Information Theory, Fall 2019

Practice problem set #3

You do **not** have to hand in these exercises, they are for your practice only.

1. **Relative entropy** Given two probability distributions P and Q on the same set $\mathcal{A}_X$, we define the *relative entropy* $D(P \| Q)$ by

$$D(P \| Q) = \sum_{x \in \mathcal{A}_X} P(x) \log(\frac{P(x)}{Q(x)}).$$

   (a) In what situation is $D(P \| Q)$ infinite?

   (b) Show that the (ordinary) entropy $H(P)$ can be written as

$$H(P) = \log|\mathcal{A}_X| - D(P \| U)$$

   where U is the uniform distribution over $\mathcal{A}_X$.

   (c) Show that $D(P \| Q) \geqslant 0$.

   (d) Show that $D(P \| Q) = 0$ if and only if $P = Q$ (and carefully distinguish the case where $P(x) = 0$ for some $x \in \mathcal{A}_X$).

2. **Optimality of the Huffman code**

   (a) For any probability distribution show that there exists an optimal code C with the following properties:
   - If $p(x) > p(y)$ then $l(C(x)) \leqslant l(C(y))$
   - The two longest codewords have the same length.
   - Two of the longest codewords differ only in the last bit and correspond to two of the least likely symbols.

   (b) (Optional:) Show that the Huffman code is optimal. (If you want, you can look up the solution to this exercise in the book 'Elements of Information Theory' by Cover and Thomas.)