Introduction to Information Theory, Fall 2020

Practice problems for exercise class #4

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

0. **Exercises from MacKay:** 5.21 (only for X^2 and X^3), 5.25

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

$$D(P \,\|\, Q) = \sum_{x \in \mathcal{A}_X} P(x) \log(\frac{P(x)}{Q(x)}). \tag{1}$$

   (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 $D(P \,\|\, Q) \geqslant 0$.
   Hint: Note that this is the same as the Gibbs inequality that we proved in class. Try proving it again to practice using Jensen's inequality.
   (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. **Using the wrong symbol code** Supose we are given two probability distributions P and Q on the same set $\mathcal{A}_X$. We know that there exists a uniquely decodable symbol code with codeword lengths $l(C(x)) = \lceil \log \frac{1}{Q(x)} \rceil$ and it has expected length satisfying

$$H(Q) \leqslant L(C, Q) < H(Q) + 1.$$

   Show that if we use this code for the 'wrong' distribution P the expected codelength will satisfy

$$H(P) + D(P \,\|\, Q) \leqslant L(C, P) < H(P) + D(P \,\|\, Q) + 1$$

   where the relative entropy is defined by Eq. (1).

3. **Optimality of the Huffman code (mathematics challenge):**

   (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) 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.)