Introduction to Information Theory, Fall 2020

Homework problem set #2 **due November 9, 2020**

Rules: Always explain your solutions carefully. Please hand in the assignment in groups on Canvas. In the werkcollege the TAs can tell you more about how this works.

1. **Entropy and Huffman codes (1 point):** Consider the following probability distribution:

x	A	B	C	D	E	F	G
P(x)	0.05	0.05	0.07	0.13	0.2	0.2	0.3

   (a) Compute $H(P)$ to one digit after the decimal point (or better). You can use a computer.
   (b) Construct a Huffman code $\mathcal{C}$ for P and compute the average length per symbol $L(\mathcal{C}, P)$.

2. **Subadditivity of the entropy (1 point):** The goal of this problem is to show that

$$H(X, Y) \leqslant H(X) + H(Y) \tag{1}$$

   for any two random variables X, Y with an *arbitrary* joint distribution $P(x, y)$.

   (a) Verify the following identity, where $P(x) = \sum_y P(x, y)$ and $P(y) = \sum_x P(x, y)$ denote the marginal distributions (as always):

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

   (b) Which inequality from class can now be used to prove Eq. (1)?
   (c) Can you interpret Eq. (1) in the context of compression?

   Optional problem: Show that Eq. (1) holds with equality precisely when X and Y are independent.

3. 🖩 **Huffman compression (1 point):**
   This week you will implement Huffman's algorithm discussed in class and above. To get started, open the Python notebook at https://colab.research.google.com/github/amsqi/iit20-homework/blob/master/02-homework.ipynb and follow the instructions.

   As last week, please submit both the notebook **and** a PDF printout, or provide a link to your solution on Colab. You can achieve the maximum score if your solution produces the correct output. We will only have a closer look at your code in case of problems.

   This programming problem may be a bit more difficult than last week's, so we will grade it gently. We also added some optional challenge problems in the notebook. Can you beat `zlib`*?* ☺