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

<table>
<thead>
<tr>
<th>x</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(x)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.13</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

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

2. **Subadditivity of entropy (1 point):** The goal of this problem is to prove the following: For any two random variables $X, Y$ with an arbitrary joint distribution $P(x, y)$, it holds that

$$H(X, Y) \leq H(X) + H(Y).$$  \hspace{1cm} (1)

This inequality is known as the subadditivity property of the entropy.

(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) Use part (a) together with an inequality you know from class to prove inequality (1).

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

3. **Huffman compression (1 point):**

This week you will implement Huffman’s algorithm. To get started, open the Python notebook at https://colab.research.google.com/github/amsqi/iit21-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`? 😊