Note: This exercise comes with an additional file for the second part.

Mathematical Basics

1. Use set theory and the definition of probability functions to show that:
   \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

2. Suppose we are interested in a test to detect a disease which affects one in
   \( \frac{1}{100000} \) people on average. A lab has developed a test which works but is
   not perfect. If a person has the disease, it will give a positive result with probability 0.97; if they do not, the test will be positive with probability 0.007. You took the test, and it gave a positive result. What is the
   probability that you actually have the disease?

3. Are \( X \) and \( Y \), as defined in the following table, independently distributed?
   How did you check?

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
<th>( p(X = x, Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.32</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.48</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Kullback Leibler Divergence and Cross-Entropy

To compare two probability distributions, Kullback Leibler divergence and Cross Entropy can be used. The Kullback Leibler divergence \( D(p||q) \) of two probability distributions \( p \) and \( q \) is defined as:

\[
D(p||q) = \sum_{x \in X} p(x) \cdot \log \left( \frac{p(x)}{q(x)} \right)
\]

The Cross Entropy \( H(p, q) \) of two discrete probability distributions \( p \) and \( q \) is defined as:

\[
H(p, q) = H(p) + D(p||q) = -\sum_{x \in X} p(x) \cdot \log (q(x))
\]
1. For each of the two given texts, determine the respective probability distribution $p$ or $q$ of characters.

2. For each of the two probability distributions $p$ or $q$, calculate the entropy $H(p)$ or $H(q)$.

3. For the two probability distributions $p$ and $q$, calculate the Kullback-Leibler divergences $D(p||q)$ and $D(q||p)$. Does it hold that $D(p||q) = D(q||p)$?

4. For the two probability distributions $p$ and $q$, calculate the cross entropies $H(p, q)$ and $H(q, p)$. How do you interpret the results?

---

**Note on Submission**  Please use PDF as a document format. If you need to compress files, use ZIP or GZIP.

If you attend the **mon. tutorial**, the **deadline** is **Friday 16th May, 23:59**. In that case send the solutions to *s9tsback@student.uni-saarland.de*.

If you attend the **wed. tutorial**, the **deadline** is **Sunday 18th May, 23:59**. In that case send the solutions to *kalofoli@ceid.upatras.gr*. 

2