Homework 2 (due 18 September)

You should turn in your solutions on paper at the start of the exercise session.

1. (a) Provide an alternative proof for Gibb’s Inequality where you use Jensen’s Inequality
(b) Prove that for any \( n \) positive numbers \( x_1, \ldots, x_n \), their geometric mean is at most their arithmetic mean:
\[
\left( \prod_{i=1}^{n} x_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

*Hint:* concavity of \( \log \) works well here, too.

2. Consider two binary-valued random variables \( X \) and \( Y \) with the following joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>( Y = 0 )</th>
<th>( Y = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X = 0 )</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>( X = 1 )</td>
<td>0</td>
<td>1/3</td>
</tr>
</tbody>
</table>

In other words, the case \( X = 1, Y = 0 \) never occurs, but the other combinations are all equally probable.

Calculate the values of
(a) \( H(X) \) and \( H(Y) \)
(b) \( H(X \mid Y) \) and \( H(Y \mid X) \)
(c) \( H(X,Y) \)
(d) \( H(Y) - H(Y \mid X) \)
(e) \( I(X;Y) \).

Optionally, you may wish to draw a graphical representation to illustrate these quantities (as on page 30 of lecture notes for 12 September, or as drawn on the blackboard during the lecture).

3. Give a proof for the chain rule of mutual information:
\[
I(X;Y_1,\ldots,Y_N) = \sum_{i=1}^{N} I(X;Y_i \mid Y_1,\ldots,Y_{i-1}).
\]

Show also that in the special case where the random variables \( Y_i \) are all independent, this simplifies to
\[
I(X;Y_1,\ldots,Y_N) = \sum_{i=1}^{N} I(X;Y_i).
\]

Continues on the next page!
4. There are 12 balls that look exactly the same, one of which is either heavier or lighter than the others. Your task is to determine which one of the balls is different, and whether it is heavier or lighter. To do this, you are given a balance. You can put any number of balls on each side of the balance, and the balance will tell you whether the sides are of equal weight or if the left or the right side is heavier. What is the fewest number of weighings you can use to solve the problem? Please describe your strategy and prove that it always works. If you can, also prove that the problem cannot be solved using fewer weighings.

5. Consider the simple Bernoulli model that generates independent random bits with \( \Pr[X_i = 1] = p \) for some fixed \( 0 \leq p \leq 1 \).

For sequence length \( n \), and some \( \epsilon > 0 \), the typical set \( A_n^\epsilon \) is defined as the set of sequences \( x_1, \ldots, x_n \) such that
\[
2^{-n(H(X)+\epsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}.
\]

What are the sequences in the typical set \( A_{15}^{0.1} \) under the Bernoulli model when \( p = 0.1 \)? How about \( p = 0.3 \), and \( p = 0.5 \)? What can you say about the sizes of these sets?

**Bonus Problem** We consider a binary symmetric channel with error rate \( p \). In other words, we transmit single bits, and each bit gets flipped with probability \( p \) independently of anything else. For error correction, we use a simple \( k \)-fold repetition code, where each bit is just repeated \( k \) times. Then an error occurs for a given bit, i.e., the recipient is unable to recover its correct value, if at least \( k/2 \) of the sent bits were flipped.

(a) Give an exact expression in terms of \( p \) and \( k \) for the probability of error in transmitting a single bit. Can you come up with an approximate formula that’s easier to understand and evaluate?

(b) Supposing \( p = 0.1 \) and we want the error probability to be at most 0.00001 %, what value of \( k \) would be sufficient? A reasonably tight estimate is a sufficient answer.

(c) Assume now that we actually want to transmit a byte (8 bits) so that all the bits are correct. What value of \( k \) would be sufficient, with other parameters as in the previous part?