Introduction to Information Theory, Fall 2020

Homework problem set #4 **due November 23, 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. **Arithmetic coding (1 point):** Consider the following language model for producing strings of length $N = 2$:

x_1	$P(x_1)$
`A`	$5/6$
`H`	$1/6$

$P(x_2\|x_1)$	$x_1 = \texttt{A}$	$x_1 = \texttt{H}$
$x_2 = \texttt{i}$	$1/3$	$1/5$
$x_2 = \texttt{y}$	$2/3$	$4/5$

For example, the probability of the string '`Hi`' according to this model would be $1/6 \times 1/5 = 1/30$. We sort the alphabet as follows: `A` $<$ `H` and `i` $<$ `y`.

   (a) Encode the string `Hi` using the arithmetic coding algorithm. Visualize all relevant intervals and indicate at which point the algorithm outputs which bits.
   (b) The bitstring `10` was produced by encoding a string x_1x_2 with the arithmetic coding algorithm. Determine x_1x_2.

2. **Symmetric channels (1 point):** Any channel $Q(y|x)$ can be seen as a matrix whose rows are indexed by y and whose columns are indexed by x. This matrix is called the *transition matrix*. A channel is called *weakly symmetric* if the columns of the transition matrix are permutations of each other and if all row sums are equal. In this case the capacity is easy to calculate:

$$C(Q) = \log \#\mathcal{A}_Y - H(c), \tag{1}$$

where c is an arbitrary column of the transition matrix.

   (a) Write down the transition matrix for the *binary symmetric channel* from class, confirm that the channel is weakly symmetric, and use (1) to compute its capacity.
   (b) Prove that the formula (1) holds for all weakly symmetric channels.

   Hint: Imitate the strategy discussed in class. Start with $I(X : Y) = H(Y) - H(Y|X)$*, then compute* $H(Y|X)$*, and finally show that the uniform input distribution maximizes* $H(Y)$*.*

3. 🖩 **Arithmetic coding (1 point):**
In this problem, you will implement the arithmetic coding algorithm discussed in class and compare its performance using different probabilistic models. To get started, open the notebook at https://colab.research.google.com/github/amsqi/iit20-homework/blob/master/04-homework.ipynb and follow the instructions.

   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.