1) **True or False** (20 points)

Copy each relation to your notebook and write true or false. Then, if it’s true, prove it. If it is false give a counterexample or prove that the opposite is true.

a) Let $X,Y$ be two random variables. Then $H(X - Y) \leq H(X|Y)$. [4 points]

b) For any finite alphabet random variables $H(X,Y,Z) - H(X,Y) \geq H(X,Z) - H(X)$. [4 p.]

c) Which of the following sequence of code-lengths are a valid binary huffman codes (can be more than one answer)? [8 points]

- 1,2,3,3
- 1,2,2,3
- 1,3,3,3
- 2,2,2,2

d) Assume a memoryless channel given by $p(y|x)$, and the capacity is given by $C = \max_{p(x)} I(X;Y)$. The capacity can be strictly increased by forming the output to be $Y_1 = f(Y)$. [4 points]

2) **Joint Entropy** (25 points) Consider $k$ different discrete random variables, named $X_1, X_2, ..., X_k$. Each random variable separately has an entropy $H(X_i)$, for $1 \leq i \leq k$.

a) What is the upper bound on the joint entropy $H(X_1, X_2, ..., X_k)$ given that $H(X_i)$, for $1 \leq i \leq k$ are fixed?

b) Under what conditions will this upper bound be reached?

c) What is the lower bound on the joint entropy $H(X_1, X_2, ..., X_k)$ given that $H(X_i)$, for $1 \leq i \leq k$ are fixed?

d) Under what condition will this lower bound be reached?

e) Assume the vector $[X_1, X_2, ..., X_k]$ is observed many times and one would like to compress it. Denote at time $i$ the observed random vector as $[X_{1,i}, X_{2,i}, ..., X_{k,i}]$. The distribution of $[X_{1,i}, X_{2,i}, ..., X_{k,i}]$ is according to $[X_1, X_2, ..., X_k]$ for all $i$ and for $i \neq j$ $[X_{1,i}, X_{2,i}, ..., X_{k,i}]$ is independent of $[X_{1,j}, X_{2,j}, ..., X_{k,j}]$.

Given that you have the possibility to optimally compress without any loss a sequence of scalar random variable distributed i.i.d. but not a sequence of random vectors. Provide a coding schemes (using the scalar compressing scheme) for an optimal lossless compression for the set of random vectors $\{[X_{1,i}, X_{2,i}, ..., X_{k,i}]\}_{i=1}^n$ where $n$ is a very large number, given that they are distributed according to the distributions of $[X_1, X_2, ..., X_k]$ that you found in subexercise 2b and subexercise 2d. (In other words, you are able to optimally compress a sequence of scalar random variable $X_i$ distributed i.i.d. , how would you use it to optimally compress the a set of vectors $[X_1, X_2, ..., X_k]$ where its distribution is according to subexercise 2b and 2d)

3) **Blahut-Arimoto’s algorithm** (35 points) Recall, that the capacity of a memoryless channel is given by

$$C = \max_{p(x)} I(X;Y).$$

Solving this optimization problem is a difficult task for the general channel. In this question we develop an iterative algorithm for finding the solution for a fixed channel $p(y|x)$.

a) (3 p.) Prove that the mutual information as a function of $p(x)$ and $p(x|y)$ may be written as

$$I(X;Y) = \sum_{x,y} p(x)p(y|x) \log \frac{p(x|y)}{p(x)}.$$

b) Show that $I(X;Y)$ as written above is concave in both $p(x)$, $p(x|y)$ (Hint. You may use the
log-sum-inequality). [8 points]

c) Find an expression for \( p(x) \) that maximizes \( I(X;Y) \) when \( p(x|y) \) is fixed (Hint. You may use the Lagrange multipliers method with the constraint \( \sum_x p(x) = 1 \). No need to take into account that \( p(x) \geq 0 \) since it will obtained anyway.) [10 points]

d) Find an expression for \( p(x|y) \) that maximizes \( I(X;Y) \) when \( p(x) \) is fixed (Hint. You may use the Lagrange multipliers method with constraints \( \sum_x p(x|y) = 1 \) for all \( y \). No need to take into account that \( p(x|y) \geq 0 \) since it will obtained anyway.). [9 points]

e) Using (d), conclude that \( C = \max_{p(x),p(x|y)} I(X;Y) \). [5 points]

The Blahut-Arimoto’s algorithm is performed by maximizing in each iteration over another variable; first over \( p(x) \) when \( p(x|y) \) is fixed, then over \( p(x|y) \) when \( p(x) \) is fixed, and so on. This iterative algorithm converges, and hence one can find the capacity of any DMC \( p(y|x) \) with reasonable alphabet size.

Reminder of Lagrange multiplier (or more generally the KKT condition): We all know that if we have a concave function \( f(x) \) with a finite maximum and its derivative exists, then the maximum is obtained at \( x \) that satisfies \( \frac{df(x)}{dx} = 0 \). Now, if we need to find the maximum of the concave function \( f(x) \) only for \( x \) that satisfies the affine condition \( h(x) = 0 \), then we define \( L(x, \nu) = f(x) + \nu h(x) \) and the maximum is obtained at \( x \) that satisfies \( \frac{dL(x,\nu)}{dx} = 0 \) and \( h(x) = 0 \); this gives us two equations that allow us to find \( x \) and \( \nu \). More generally, in case that \( \overline{x} = [x_1, x_2, ..., x_k] \) is a vector of size \( k \) and we have \( l \) affine constraints, then \( L(\overline{x}, \overline{\nu}) = f(\overline{x}) + \sum_{j=1}^{l} \nu_j h_j(\overline{x}) \) and the maximum is obtained at \( \overline{x} \) that satisfies \( \frac{dL(\overline{x},\overline{\nu})}{d\overline{x}_i} = 0 \) for all \( 1 \leq i \leq k \) and \( h_j(\overline{x}) = 0 \) for all \( 1 \leq j \leq l \).

4) **Source-channel coding problem** (20 points) Consider the source-channel coding problem given in Fig. 1, where \( V, X, Y, \hat{V} \) have a Binary alphabet. The source \( V \) is i.i.d. Bernoulli \( (p) \), and the channel is in Fig. 2.

![Fig. 1. A source-channel coding problem](image1)

![Fig. 2. The channel.](image2)

- a) Assume that error-free bits can be transmitted through the channel. What is the minimum rate in which the source \( V \) should be encoded such that the source decoder can reconstruct the source \( V \) losslessly? [4 points]
- b) What is the capacity of the channel given in Fig. 2? [8 points]
- c) For what values of \( p \) can the source \( \hat{V} \) be reconstructed losslessly using the scheme in Fig. 1 (you may use the inverse of \( H \), i.e., \( H^{-1}(q) \))? [4 points]
- d) Would the answer to 4c changes if a joint source-channel coding and decoding is allowed?[4 p.]

Good Luck!