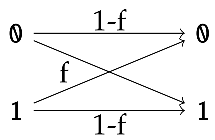
Introduction to Information Theory, Fall 2020

Practice problems for exercise class #8

You do **not** have to hand in these exercises, they are for your practice only.

1. **Binary symmetric channel:** Consider the binary symmetric channel Q which transmits a given bit X correctly with probability $1 - f$ and flips the bit with probability f:



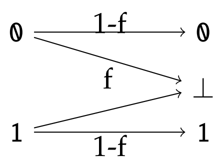
Let Y denote the output of the channel. Some of you discussed the capacity of this channel already in the lecture, but here you have to do the chance to do it yourself again to make sure you understand all steps.

(a) Write down the transition probabilities $Q(y|x)$ for the binary symmetric channel.
(b) Compute the mutual information $I(X : Y)$ for the input distribution of X with $P(0) = P(1) = 1/2$ and $f = 1/4$.
(c) Compute $H(Y|X)$ for an arbitrary input distribution and show that it does not depend on the input distribution.
(d) Show that the capacity of the binary symmetric channel is given by

$$C(Q) = 1 - H(\{f, 1 - f\}).$$

What is the optimal input distribution? Why is this intuitively sensible? *Hint: write* $I(X : Y) = H(Y) - H(Y|X)$ *and use your result from (c).*

2. **Binary erasure channel:** Consider the binary erasure channel Q which transmits a given bit X correctly with probability $1 - f$ and gives an error message $\perp$ with probability f:



Let Y denote the output of the channel.

(a) Write down the transition probabilities $Q(y|x)$ for the binary erasure channel.
(b) Compute the mutual information $I(X : Y)$ for the input distribution of X with $P(0) = P(1) = 1/2$ and $f = 1/4$.
(c) Compute $H(X|Y)$ for an arbitrary input distribution.
(d) Compute the capacity of the binary erasure channel.

3. **Markov chains and data processing:** Suppose we are given three (correlated) random variable X, Y and Z. Then we can always write

$$P(x, y, z) = P(x)P(y|x)P(z|x, y).$$

If we can actually write

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

then we say that $X \to Y \to Z$ forms a *Markov chain*, which means essentially that Y depends on X, and Z depends on Y but not on X.

(a) Show that $X \to Y \to Z$ is a Markov chain if and only if $P(x, z|y) = P(x|y)P(z|y)$. Argue that if $X \to Y \to Z$ is a Markov chain, then $Z \to Y \to X$ is also a Markov chain.

(b) Show that the following inequality holds for any three random variables (whether they form a Markov chain or not):

$$H(Z|X, Y) \leqslant H(Z|Y)$$

Moreover, show that equality holds if and only if $X \to Y \to Z$ is a Markov chain.

(c) Prove the *Data Processing Inequality*: if $X \to Y \to Z$ is a Markov chain then

$$H(X|Y) \leqslant H(X|Z)$$

and

$$I(X : Y) \geqslant I(X : Z).$$

Explain what this tells you about data processing.