TUTORIAL II

0 Homework 1

1. (Repetition code) Suppose that you have a disk drive where each bit gets flipped with probability $f = 0.1$ in a year. In order to be able to correct errors, we take a copy of the full drive $N - 1$ times so that we have N copies of the original data (N is odd). After one year, I would like to retrieve a given bit of the original drive. What should I do? Suppose I want the probability of error for this bit to be at most δ, how large should I take N as a function of δ? How large is this for $\delta = 10^{-10}$?

2. Let $X \in \mathbb{N}$ be a discrete random variable and $g : \mathbb{N} \to \mathbb{N}$. What can you say in general on the relation between $H(X)$ and $H(g(X))$? And in particular, if $g(n) = 2^n$?

1 A realistic find query

We consider a list of 32 elements and we want to test if a given element z belongs to the list or not. We assume that the probability that the element belongs to the list is $1/2$, and that all the positions within the list are equiprobable. Our strategy is to test the first element, then the second element, ... until the wanted element is found or the end of the list is reached. We denote by F the random variable which is equal to 1 if and only if z is in the list, 0 otherwise.

1. Compute the entropy of F.

2. We denote by L_1 the random variable corresponding to the result of the first test. Compute the entropy of L_1.

3. Compute the distribution of the joint variable (F, L_1), and give the joint entropy $H(F, L_1)$.

4. Compute the conditional entropy $H(F|L_1)$.

5. We denote by $L_1, \dots, L_n$ the result of the successive tests. Compute directly the conditional entropy $H(F|L_1, \dots, L_n)$.

6. If we plug $n = 16$ in the previous solution, we find $0.689 > \frac{1}{2}$. Is it reasonable ? What is the value of $H(F|L_1, \dots, L_{32})$?

2 Data processing inequality for mutual information

Recall that:

$$H(X|Y) \stackrel{\text{def}}{=} \sum_{y \in A_Y} P_Y(y) H(X|Y = y) \quad , \quad H(X,Y) = H(X) + H(Y|X) \quad \text{and} \quad I(X;Y) \stackrel{\text{def}}{=} H(X) - H(X|Y)$$

0. We know that more information cannot increase uncertainty in the sense that $H(X|Y) \leq H(X)$. Show that this is not true if we do not take the average of Y, i.e. give an example of a pair of random variables (X, Y) such that $H(X|Y = y) > H(X)$ for some y.

We define the conditional mutual information:

$$I(X;Y|Z) \overset{\text{def}}{=} H(X|Z) - H(X|Y,Z)$$

If X and Z are conditionally independent given Y (i.e. $\mathbf{P}_{Z|Y,X} = \mathbf{P}_{Z|Y}$), we will use the notation $X \to Y \to Z$ (this notation is motivated by the theory of Markov chains). Notice that $X \to Y \to Z$ implies $Z \to Y \to X$ since $\mathbf{P}_{Z|Y,X} = \mathbf{P}_{Z|Y} \Rightarrow \mathbf{P}_{X|Y,Z} = \mathbf{P}_{X|Y}$.

1. Show that $I(X;Y|Z)$ is the average over Z of $I(X;Y)$, i.e.: $I(X;Y|Z) = \sum_z \mathbf{P}(Z = z) I(X|Z = z; Y|Z = z)$.

2. Show that $I(X;(Y,Z)) = I(X;Z) + I(X;Y|Z)$

3. For any $X \to Y \to Z$, show that the conditional mutual information $I(X;Z|Y)$ is 0.

4. Using question 2 and 3, show the data processing inequality: $I(X;Y) \geq I(X;Z)$ for any $X \to Y \to Z$.

5. Show that for any function g, we have $I(X;Y) \geq I(X;g(Y))$.

3 A general compressor for variable-length lossless compression

Recall that in variable length lossless compression, we aim to optimize the expected length $\mathbb{E}(|C(X)|)$. For this, define an optimal compressor as follows: we order all the possible bitstrings $0, 1$ in the shortlex order, i.e., $w \leq_{slex} w'$ if $|w| < |w'|$, or $|w| = |w'|$ and $w \leq_{lex} w'$; and let w_i the i-th bitstring in this order. For example, $w_1 = \emptyset$ is the empty string, $w_2 = 0$, $w_3 = 1$, $w_4 = 00$, etc...

Let P_X be a distribution and $x_1, \ldots, x_{|\mathcal{X}|}$ be such that $P_X(x_1) \geq \cdots \geq P_X(x_{|\mathcal{X}|})$.

1. Define C^* as $C^*(x_i) = w_i$ for $i = 1, \cdots, |\mathcal{X}|$. Show that it is an optimal compressor, i.e., $\mathbb{E}\{|C^*(X)|\} \leq \mathbb{E}\{|C(X)|\}$ for any lossless compressor C.

2. Prove that $\mathbb{E}\{|C^*(X)|\} \leq H(X)$

3. Prove that $H(X) - \log_2(1 + \lfloor \log_2 |\mathcal{X}| \rfloor) \leq \mathbb{E}\{|C^*(X)|\}$