TD 1

Exercise 1. In the Death Valley:

- it rains on average once in 100 days;
- the weather forecast foresees 3 rainy days in 100 days;
- each time it rains, the weather forecast is right;
- Guy-who-knows-everything predicts that it never rains.

Who is better, Guy-who-knows-everything or the weather forecast ?

Exercise 2. Let $(X_1, X_2, \dots, X_N)$ be an N-tuple of random variables. Its entropy is defined as

$$H(X_1 X_2 \dots X_N) = -\sum p(x_1, \dots, x_N) \log p(x_1, \dots, x_N)$$

1. (independent r.v.) Show that if $X_1, \dots, X_n$ are independent, then

$$H(X_1 X_2 \dots X_N) = \sum_{i=1}^{N} H(X_i)$$

2. (general case) Show the "chain rule for entropy"

$$H(X_1 X_2 \dots X_N) = H(X_N | X_1, \dots X_{N-1}) + H(X_{N-1} | X_1 \dots X_{N-2}) + \dots + H(X_2 | X_1) + H(X_1)$$

Exercise 3. Let (X, Y, Z) be a tuple of random variables such that

$$\begin{aligned} p_{XYZ}(0,0,0) &= \frac{1}{4} \\ p_{XYZ}(0,1,0) &= \frac{1}{4} \\ p_{XYZ}(1,0,0) &= \frac{1}{4} \\ p_{XYZ}(1,0,1) &= \frac{1}{4} \end{aligned}$$

Recall that

$$H(Y|X) = \sum_x p(x) H(Y|X = x)$$

Compute $H(X)$, $H(Y|X)$, $H(Z|X,Y)$. Find $H(X,Y,Z)$ in two different manners (chain rule and direct computation). Recall that $h(1/4) \approx 0.811$. Compute $H(Y)$ and verify that $H(Y|X) \leq H(Y)$. How much information conveys X about Y and how much information conveys Y about X ?

Exercise 4. Show that $H(Y|X) = 0$ if and only if Y is a function of X, that is for all x, such that $p(x) > 0$ there exists y such that $p(y|x) = 1$. We use here the notation $p(y|x) = p(Y = y|X = x)$.

Exercise 5. Let X be a random variable and $g(x)$ be a function. By using the chain rule in two different ways, show that

$$H(g(X)) \leq H(X).$$

Exercise 6. Let X be a random variable with entropy $H(X)$. What is the relationship between $H(Y)$ and $H(X)$

- when $Y = 2^X$
- when $Y = \cos X$.

Exercise 7. We assume here that we have a twin-pan scale (which only allows to compare two objects that have been put on the pans) and n coins. A genuine coin has a certain weight w which is unknown. A counterfeit coin has a different weight, however this weight can be either smaller or larger than the weight of a genuine coin. We know that that among these n coins there is at most one counterfeit coin.

1. Give the best lower bound you can find on the number of weighings that have to be performed to detect with certainty a counterfeit coin. *Hint: use Exercise 5...*
2. Assume now that we also have an auxiliary infinite stack of genuine coins. Give the optimal strategy to find the counterfeit coin if there is one when $n = 13$. Generalize your strategy to $n = \frac{3^k-1}{2}$.

Exercise 8 (The password problem)**.** A hacker wants to access a server which is protected by a password that is unknown to the hacker. We assume that the passwords are binary sequences of length m and we let $\mathcal{M} = \{0,1\}^m$ be the set of all possible passwords. We also assume that the only way for the hacker to access the server is find the right password by checking them one by one.

It turns out that the password is chosen randomly in $\mathcal{M}$ according to a probability distribution of entropy $h \leq m$. We denote by p_i the probability of the i-th element of $\mathcal{M}$ and we assume that the order on $\mathcal{M}$ is chosen in such a way that

$$p_1 \geq p_2 \geq p_3 \cdots \geq p_{2^m}.$$

1. Show that the best strategy consists in trying the elements according to the aforementioned order, i.e. try first the element of index 1, that is the element of probability p_1, then the second element and so on and so forth. Give a formula for $\mathcal{N}(p)$ which is the expected number of passwords that have to be tested before finding the right one.
2. Consider now two probability distributions over the positive integers $p = (p_i)_{i\geq 1}$ and $q = (q_i)_{i\geq 1}$ such that the p_i's and the q_i's are decreasing and $q_i > 0$ for all i. On the other hand p_i is allowed to be equal to zero for i large enough. Recall that the Kullback distance from p to q is given by
$$D(p \parallel q) = \sum_{i\geq 1} p_i \log \frac{p_i}{q_i}. \tag{1}$$
Let $q_i = (1-\alpha)\alpha^{i-1}$ for a certain real $0 < \alpha < 1$. Show that if $H(p) = H(q)$, then $\sum_{i\geq 1} ip_i \geq \sum_{i\geq 1} iq_i$.
Indication: use that $D(p \parallel q) \geq 0$.
3. Express $H(q)$ in terms of α. Let H_α be this expression. We recall that $\sum_{i\geq 1} \alpha^{i-1} = 1/(1-\alpha)$ and $\sum_{i\geq 1} i\alpha^{i-1} = 1/(1-\alpha)^2$
4. Deduce that for all $0 < \alpha < 1$ we have $1 < (1-\alpha)2^{H_\alpha} < e$, where e is the base of the natural logarithm.
5. Deduce from the previous result that $\mathcal{N}(p) > c_1 2^h$ (give c_1 explicitly here). Give an interpretation of this result.

Exercise 9. Show that

$$\binom{n}{t} \leq 2^{nh(t/n)}$$

Exercise 10. [Fano lemma- relationship between error probability of an estimator and conditional probability] Let X and Y be two random variables (with X taking its values in an alphabet of size a). Let $\hat{X}$ be an estimator for X computed from the knowledge of Y (it is therefore a function of Y). Denote by P_e the error probability of the estimator, that is $P_e = p(\hat{X} \neq X)$. Show that

$$h(P_e) + P_e \log_2(a-1) \geq H(X|Y).$$

Hint : introduce a random variable E defined by

$$E = \begin{cases} 1 & \text{si } \hat{X} \neq X \\ 0 & \text{si } \hat{X} = X \end{cases}$$

and find two different expressions for $H(E, X|Y)$. Prove that Fano's inequality is sharp by considering the following example. X takes a different values $1, \dots, a$ with probability $p_i \overset{\text{def}}{=} \mathbf{Prob}(X = i)$ and

$$p_1 \geq p_2 = \dots = p_a$$

and Y is independent of X.

Exercise 11. Let X and Y be two random variables taking their values in a group $(G, +)$. Let $Z = X + Y$.

1. Show that $H(Z|X) = H(Y|X)$.

2. Show that if X et Y are independent $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$ (use non-negativity of mutual information).

3. Give an example of two random variables X and Y such that $H(X) > H(Z)$ and $H(Y) > H(Z)$.