EXERCISES 1

Excercise 1 (The relative entropy)**.** We define the *relative entropy* of probability distributions $p, q \in \mathcal{P}(\{1, \dots, d\})$ as

$$D(p\|q) = \begin{cases} \sum_{i=1}^d p_i \log_2\left(\frac{p_i}{q_i}\right) & \text{, if } \mathrm{supp}(p) \subseteq \mathrm{supp}(q) \\ +\infty & \text{, otherwise.} \end{cases}$$

The relative entropy measures statistical difference and it is closely related to other entropic quantities. In this exercise, you will prove some important properties of the relative entropy and use them to determine properties of the Shannon entropy and the mutual information.

(1) For positive numbers $a_1, \dots, a_d \in \mathbb{R}^+$ and $b_1, \dots, b_d \in \mathbb{R}^+$ show that

$$\sum_{i=1}^d a_i \log_2\left(\frac{a_i}{b_i}\right) \geqslant \left(\sum_{i=1}^d a_i\right) \log_2\left(\frac{\sum_{i=1}^d a_i}{\sum_{i=1}^d b_i}\right).$$

Hint: Reformulate this inequality using the variables

$$\beta_i = \frac{b_i}{\sum_{j=1}^d b_j} \quad \text{and} \quad t_i = \frac{a_i}{b_i}.$$

(2) Show that

$$D(p\|q) \geqslant 0,$$

for all $p, q \in \mathcal{P}(\{1, \dots, d\})$ with equality if and only if $p = q$.

(3) Given $p_1, p_2, q_1, q_2 \in \mathcal{P}(\{1, \dots, d\})$ and $\lambda \in [0, 1]$ show that

$$D((1-\lambda)p_1 + \lambda p_2 \| (1-\lambda)q_1 + \lambda q_2) \leqslant (1-\lambda) D(p_1\|q_1) + \lambda D(p_2\|q_2).$$

(4) Show that the Shannon entropy can be written as

$$H(p) = \log_2(d) - D(p\|u),$$

for the uniform distribution $u \in \mathcal{P}(\{1, \dots, d\})$. Conclude that $H(p) \leqslant \log_2(d)$ with equality if and only if $p = u$, and that $p \mapsto H$ is concave.

(5) Consider a joint distribution $p_{AB} \in \mathcal{P}(\{1, \dots, d\} \times \{1, \dots, d\})$ and define the distribution $p_A \times p_B$ by $(p_A \times p_B)(i, j) = p_A(i) p_B(j)$. Show that the mutual information can be written as

$$I(A : B)_{p_{AB}} = D(p_{AB} \| p_A \times p_B).$$

Conclude that the mutual information is positive and that the Shannon capacity is convex in the communication channel.

(6) Consider finite alphabets Σ_A and Σ_B. Which channels $N : \Sigma_A \to \mathcal{P}(\Sigma_B)$ have zero capacity?

Excercise 2 (Capacity of binary symmetric channels)**.** Consider a communication channel $N_p : \{0, 1\} \to \mathcal{P}(\{0, 1\})$ given by

$$N_p(0) = \begin{pmatrix} 1-p \\ p \end{pmatrix} \quad \text{and} \quad N_p(1) = \begin{pmatrix} p \\ 1-p \end{pmatrix},$$

for some $p \in [0, 1]$. Such a channel is called *binary* and *symmetric* since it acts on a binary alphabet and flips the any of the possible input bits with the same probability p (see Figure 1). Use the capacity formula to show that

$$C(N_p) = 1 - h_2(p),$$

where $h_2 : [0, 1] \to \mathbb{R}$ is the binary entropy function given by $h_2(p) = -p \log_2(p) - (1 - p) \log_2(1 - p)$.

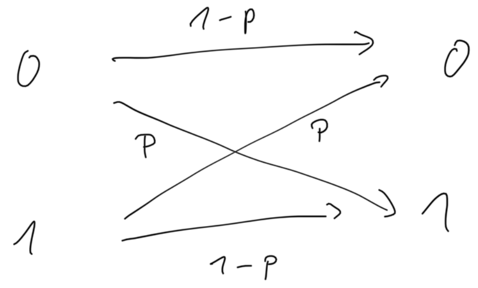

Figure 1. Binary symmetric channel N_p

Excercise 3 (The statistical distance)**.** Consider the following scenario: Alice and Bob conduct an experiment. Alice has a machine with two buttons labeled "0" and "1". After pressing the button "0" the machine emits a classical system in the probabilistic state $p_0 \in \mathcal{P}(\{1, \ldots, n\})$, and after pressing the button "1" the machine emits a classical system in the probabilistic state $p_1 \in \mathcal{P}(\{1, \ldots, n\})$. Bob catches the classical system, and by looking at it he tries to guess which button Alice pressed. Assume that Alice presses the button "0" with probability $\lambda \in [0, 1]$ and the button "1" with probability $1 - \lambda$, and that Bob knows the distributions p_0, p_1 and λ. What is the optimal success probability with which Bob can guess which button Alice pressed?

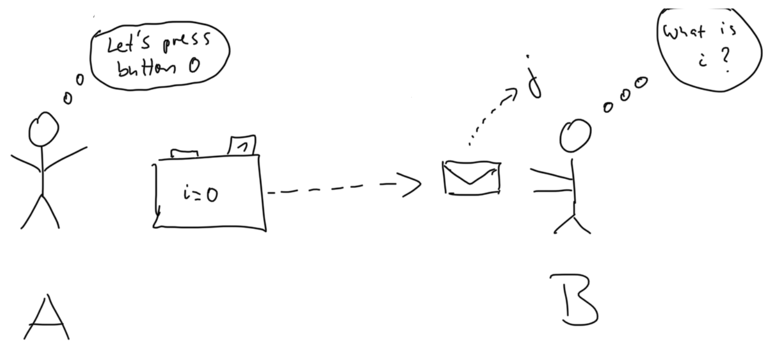

Figure 2. Alice and Bob doing experiments.

(1) Use Bayes' theorem to compute the conditional probabilities $P(A = i|B = j)$ of Alice pressing button i and Bob observing the value j.
(2) Consider the following strategy: After observing the value j, Bob guesses that Alice pressed the button i maximizing the conditional probability $P(A = i|B = j)$. Observe that this strategy is the optimal deterministic strategy (i.e, assigning a fixed guess to each of the possible j Bob could observe).

(3) Show that the success probability of this strategy is given by

$$p_{\text{opt}} = \frac{1}{2} + \frac{1}{2}\|\lambda p_0 - (1-\lambda)p_1\|_1,$$

with the *statistical distance*

$$\|x - y\|_1 = \sum_{i=1}^{n} |x_i - y_i|.$$

Hint: Consider first the difference between the probabilities of guessing correctly and incorrectly.

(4) Could Bob achieve a better success probability than p_{opt} from above, if he uses a probabilistic strategy (i.e., where he determines his guess from the observed value j using some random process)?

Excercise 4 (Asymptotic discrimination of probability distributions)**.** Now, we consider the same scenario as in the previous exercise involving the two characters Alice and Bob. This time Alice presses the same button multiple times in a row. How does this change the optimal probability for Bob guessing which button Alice has pressed? To answer this question, we proceed as follows:

(1) Show that

$$\|\lambda p - (1-\lambda)q\|_1 \geqslant 1 - 2\sqrt{\lambda(1-\lambda)}\sum_{i=1}^{n}\sqrt{p_i q_i},$$

for any pair of distributions $p, q \in \mathcal{P}\{1, \ldots, n\}$ and any $\lambda \in [0,1]$.

(2) Now, show that $p \neq q$ if and only if

$$\|\lambda p^{\times n} - (1-\lambda)q^{\times n}\|_1 \to 1,$$

as $n \to \infty$.

(3) What does this mean for the success probability in the asymptotic scenario where Bob tries to guess which button Alice pressed repeatedly?