EXERCISES 1

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

$$D(p\|q) = \begin{cases} \sum_{x\in\Sigma} p(x) \log\left(\frac{p(x)}{q(x)}\right) & , \text{ if } \operatorname{supp}(p) \subseteq \operatorname{supp}(q) \\ +\infty & , \text{ otherwise.} \end{cases}$$

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 $a_1, \ldots, a_d \in \mathbb{R}^+$ and $b_1, \ldots, b_d \in \mathbb{R}^+$ show that

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

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 Gibb's inequality

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

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

(3) Given $p_1, p_2, q_1, q_2 \in \mathcal{P}(\Sigma)$ and $\lambda \in [0, 1]$ show that the relative entropy is jointly convex, i.e.,

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

(4) Show that the Shannon entropy of a distribution $p \in \mathcal{P}(\Sigma)$ can be written as

$$H(p) = \log(|\Sigma|) - D(p\|u),$$

for the uniform distribution $u \in \mathcal{P}(\Sigma)$, i.e., the distribution with $u(x) = 1/|\Sigma|$. Conclude that $H(p) \leqslant \log(|\Sigma|)$ with equality if and only if $p = u$, and that the entropy function is concave.

Excercise 2 (Conditional entropy and mutual information)**.** Let $p_{AB} \in \mathcal{P}(\Sigma_A \times \Sigma_B)$ denote a joint probability distribution and (X, Y) a pair of random variables distributed according to p_{AB}. We define the conditional entropy of X given Y as

$$H(X|Y) = \sum_{(x,y)\in\Sigma_A\times\Sigma_B} p_{AB}(x,y) \log\left(\frac{1}{p_{AB}(x|y)}\right).$$

The mutual information is given by

$$I(X:Y) = H(X) - H(X|Y),$$

where we write $H(X)$ for the marginal entropy, i.e., the entropy $H(p_A)$ of the marginal distribution of X.

(1) Show that the mutual information is symmetric, i.e., that

$$I(X:Y) = I(Y:X).$$

(2) Show that the mutual information can be written as

$$I(X:Y) = D\left(p_{AB} \| p_A \times p_B\right).$$

Conclude that $I(X:Y) \geqslant 0$ and that $I(X:Y) = 0$ holds if and only if X and Y are independent.

(3) Show that conditioning cannot increase the entropy, i.e., that $H(X) \geqslant H(X|Y)$. When do we have equality here?

Excercise 3 (Joint entropies)**.** Let $p_{AB} \in \mathcal{P}(\Sigma_A \times \Sigma_B)$ denote a joint probability distribution and (X, Y) a pair of random variables distributed according to p_{AB}. The joint entropy of (X, Y) is defined as

$$H(X,Y) = H(p_{AB}).$$

(1) Prove the so-called chain rule of the entropy

$$H(X,Y) = H(X) + H(Y|X).$$

(2) Prove that subadditivity of the entropy, i.e., the inequality

$$H(X,Y) \leqslant H(X) + H(Y),$$

and that we have equality if and only if X and Y are independent.

(3) Consider the joint probability distribution $p_{AB} \in \mathcal{P}(\{1,2,3,4\} \times \{1,2,3,4\})$ given by

$$(p_{AB}(i,j))_{i,j} = \begin{pmatrix} 1/8 & 1/16 & 1/32 & 1/32 \\ 1/16 & 1/8 & 1/32 & 1/32 \\ 1/16 & 1/16 & 1/16 & 1/16 \\ 1/4 & 0 & 0 & 0 \end{pmatrix},$$

and a pair of random variables (X, Y) distributed according to p_{AB}. Compute the joint entropy $H(X,Y)$ the marginal entropies $H(X)$ and $H(Y)$, the conditional entropy $H(X|Y)$ and the mutual information $I(X:Y)$.

Excercise 4 (Asymptotic discrimination of probability distributions)**.** Recall the experiment discussed in the lecture: Alice has a machine with two buttons labeled "0" and "1". After pressing a button the machine emits a letter containing a card with a single letter $x \in \Sigma$ from some finite alphabet Σ. If Alice presses the button "0", then the letter will be chosen according to a probability distribution $p_0 \in \mathcal{P}(\Sigma)$, and if she presses the button "1", then it will be chosen according to a probability distribution $p_1 \in \mathcal{P}(\Sigma)$. Bob now catches the letter, looks at the card inside, and tries to guess which button Alice has pressed. If we 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.

Now, we consider the same scenario 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

$$\|\lambda p - (1-\lambda)q\|_1 \geqslant 1 - 2\sqrt{\lambda(1-\lambda)} \sum x \in \Sigma \sqrt{p(x)q(x)},$$

for any pair of distributions $p, q \in \mathcal{P}(\Sigma)$ 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?