# Chapter 5

# Information Theory

## Exercise 1

Let us consider a random variable $X$ given by

$$X = \begin{cases} a & \text{with probability } 1/2 \\ b & \text{with probability } 1/4 \\ c & \text{with probability } 1/8 \\ d & \text{with probability } 1/8 \end{cases}.$$

Calculate entropy of $X$.

**Answer of exercise 1**

The entropy of $X$ is

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}.$$

## Exercise 2

Let $(X, Y)$ have the following joint distribution:

| | $X=1$ | $X=2$ | $X=3$ | $X=4$ |
|---|---|---|---|---|
| $Y=1$ | $\frac{1}{8}$ | $\frac{1}{16}$ | $\frac{1}{32}$ | $\frac{1}{32}$ |
| $Y=2$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{1}{32}$ | $\frac{1}{32}$ |
| $Y=3$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ |
| $Y=4$ | $\frac{1}{4}$ | $0$ | $0$ | $0$ |

Compute $H(X), H(Y)$ and $H(X|Y)$.

**Answer of exercise 2**

The marginal distribution of $X$ is $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$ and the marginal distribution of $X$ is $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, hence $H(X) = \frac{7}{4}$ bits and $H(Y) = 2$ bits.

$$\begin{aligned}
H(X|Y) &= \sum_{i=1}^{4} p(Y=i) H(X|Y=i) \\
&= \frac{1}{4} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right) + \\
&\quad \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{4} H(1,0,0,0) \\
&= \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 0 \\
&= \frac{11}{8}.
\end{aligned}$$

## Exercise 3

Let $A = \{0, 1\}$ and consider two distributions $p$ and $q$ on $A$. Let $p(0) = 1 - r, p(1) = r$, and let $q(0) = 1 - s$, and $q(1) = s$. Compute $D(p \parallel q)$ and $D(q \parallel p)$ for $r = s$ and for $r = \frac{1}{2}, s = \frac{1}{4}$.

**Answer of exercise 3**

We have that

$$D(p \parallel q) = (1-r) \log \frac{1-r}{1-s} + r \log \frac{r}{s}$$

and

$$D(q \parallel p) = (1-s) \log \frac{1-s}{1-r} + s \log \frac{s}{r}.$$

If $r = s$, then $D(p \parallel q) = D(q \parallel p) = 0$.

For $r = \frac{1}{2}, s = \frac{1}{4}$, we calculate

$$D(p \parallel q) = \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{3}{4}} + \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{1}{4}} = 1 - \frac{1}{2} \log 3 = 0.2075$$

and

$$D(q \parallel p) = \frac{3}{4} \log \frac{\frac{3}{4}}{\frac{1}{2}} + \frac{1}{4} \log \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{3}{4} \log 3 - 1 = 0.1887.$$

## Exercise 4

Let $(X, Y)$ have the same distribution as in Exercise 2. Compute their mutual information $I(X;Y)$.

**Answer of exercise 4**

We can rewrite definition of mutual information $I(X;Y)$ as

$$\begin{aligned}
I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} \\
&= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x|y) \\
&= -\sum_{x} p(x) \log p(x) - \left( -\sum_{x,y} p(x,y) \log p(x|y) \right) \\
&= H(X) + H(X|Y)
\end{aligned}$$

From the results of Exercise 2, we get

$$I(X;Y) = \frac{7}{4} - \frac{11}{8} = \frac{3}{8} = 0.375.$$

## Exercise 5

Let $(X,Y)$ have the following joint distribution:

| | $X=1$ | $X=2$ |
|---|---|---|
| $Y=1$ | $0$ | $\frac{3}{4}$ |
| $Y=2$ | $\frac{1}{8}$ | $\frac{1}{8}$ |

Compute $H(X), H(X|Y=1), H(X|Y=2)$ and $H(X|Y)$.

**Answer of exercise 5**

By calculation,

$$\begin{aligned}
H(X) &= H\left(\frac{1}{8}, \frac{7}{8}\right) = 0.544 \\
H(X|Y=1) &= 0 \\
H(X|Y=2) &= 1 \\
H(X|Y) &= \frac{3}{4} H(X|Y=1) + \frac{1}{4} H(X|Y=2) = 0.25
\end{aligned}$$

Thus the uncertainty about $X$ is increased if $Y=2$ is observed and decreased if $Y=1$, but uncertainty decreases on the average.

## Exercise 6

Find random variables $X$, $Y$ and $y \in \mathcal{Y}$ such that $H(X) < H(X \mid Y=y)$.

## Exercise 7

What is the minimum entropy for $H(p_1, p_2, \ldots, p_n) = H(\vec{p})$ as $\vec{p}$ ranges over all probability vectors? Find all possible values of $\vec{p}$ which achieve this minimum.

## Exercise 8

What is the general inequality relation between $H(X)$ and $H(Y)$ if

1. $Y = 2^X$
2. $Y = \cos X$

## Exercise 9

Show that whenever $H(Y \mid X) = 0$, $Y$ is a function of $X$.

## Exercise 10

A metric $\rho$ on a set $X$ is a function $\rho\colon X \times X \to \mathbb{R}$. For all $x, y, z \in X$, this function is required to satisfy the following conditions:

1. $\rho(x, y) \geq 0$
2. $\rho(x, y) = 0$ if and only if $x = y$
3. $\rho(x, y) = \rho(y, x)$
4. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$

For the metric $\rho(X, Y) = H(X|Y) + H(Y|X)$, show that conditions 1, 3 and 4 hold. Should we define $X = Y$ iff there exists a bijection $f$ such that $X = f(y)$, show that 2 holds as well.