Solution to exercise sheet n° 1 :

1-1.

(a) Example $\mathcal{X} = \{0, 1, \dots m-1\}$. With $p(x) = \frac{1}{m}$ for $x \in \mathcal{X}$.

(b) $H(X) = -\sum_{i=0}^{m-1} p_i \log_2 p_i = \log_2 m = 6$ bits.

(c) One needs 6 bits, since $2^6 = 64$.

(d) One needs 3 symbols, since $4^3 = 64$.

(e) One defines the Lagrangian $L(\{p_i\}) = H(X) - \lambda[\sum_{i=0}^{m-1} p_i - 1]$.
Condition for an extremum :

$$\forall i, \frac{\partial L}{\partial p_i} = 0.$$

The distribution that maximizes $H(X)$ (note that $H(X)$ is concave) satisfies :

$$-\log_2 p_i - \frac{1}{\ln 2} + \lambda = 0 \qquad \forall i.$$

One finds $p_i = 2^{\lambda - \frac{1}{\ln 2}}$, *i.e.* p_i is a constant. With the constraint one finds $p_i = \frac{1}{m}$.

1-2.

(a) $H(X_P) = 1.75$ bits.

(b) $H(X_Q) = 2$ bits.

(c) The expected length of the codewords is 1.75 bits for the distribution P and 2.25 bits for the distribution Q.

(d) The entropy gives the minimal expected length of codewords one can obtain. The binary code C is optimal for the distribution P, since its expected length $L_P = H(X_P)$. For the distribution Q we find $L_Q > H(X_Q)$ and $L_Q > L_P$, which implies that the code is not optimal. The optimal code for Q is given by a simple enumeration of the elements of X ; it is thus, impossible to compress that source.

1-3.

(a) $H(X) = 2$ bits.

(b) Sequence of questions :
Did one obtain "head" with the first flip ?
Did one obtain "head" with the second flip ?
...
Did one obtain "head" with the nth flip ?
One can associate the answer to each question with a separate bit, and thus the answers to n questions are encoded in n bits. We obtain the expected number of "yes/no" questions : $\sum_{n=1}^{\infty} p(n)n = H(X) = 2$. It is equal to the entropy, which shows that the sequence of questions is optimal.

1-4.

(a) $H(Y) = H(X) = 1.875$ bits, because the function is bijective (i.e. if one fixes Y, one knows X).

(b) The function is not bijective. One obtains $H(Y) < H(X)$ with $H(X) = 2.085$ bits and $H(Y) = 1,325$ bits.

(c) $H(X, f(X)) = H(X) + H(f(X)|X)$ but $H(f(X)|X) = 0$, because knowing X fixes $f(X)$.
$H(f(X), X) = H(f(X)) + H(X|f(X))$ but $H(X|f(X)) \geq 0$.
Finally : $H(f(X), X) = H(X, f(X))$ implies $H(f(X)) \leq H(X)$.
It is saturated if $H(X|f(X)) = 0$, i.e. if the function $Y = f(X)$ is bijective.

1-5.

(a) Definition of the conditional entropy : $H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X = x)$.

$$H(Z|Y) = \sum_{y\in\mathcal{Y}} p(y)H(Z|Y = y) = \sum_{y\in\mathcal{Y}} p(y)H(X+Y|Y = y) = \sum_{y\in\mathcal{Y}} p(y)H(X|Y = y) = H(X|Y).$$

If X and Y are independent, then $H(X|Y) = H(X)$.
As conditioning can only reduce the entropy : $H(Z|Y) \leq H(Z)$.
We finally obtain $H(X) \leq H(Z)$, and similarly $H(Y) \leq H(Z)$.

(b) Example

P($X = x, Y = y$)		$Y =-1$	-2	-3	-4	P($X = x$)
$X =$	1	1/4	0	0	0	1/4
	2	0	1/4	0	0	1/4
	3	0	0	1/8	1/8	1/4
	4	0	0	1/8	1/8	1/4
	P($Y = y$)	1/4	1/4	1/4	1/4	1

Calculation of $H(X)$ and $H(Y)$: $H(Y) = H(X) = H(1/4, 1/4, 1/4, 1/4) = \log_2 4 = 2$ bits.
One obtains $\mathcal{Z}$ $=\{3, 2, 1, 0, -1, -2, -3\}$ with $P(Z = 0) = 3/4$, $P(Z = 1) = 1/8$ et $P(Z = -1) = 1/8$ and the other probabilities are zero.
Calculation of $H(Z)$: $H(Z) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{8} = 1.061$ bits.
One verifies that $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) We require that X and Y are independent and that all $z_{i,j} = x_i + y_j$ are distinct for any couple (i, j). If these conditions are satisfied one obtains $p_z(i, j) = p_x(i)p_y(j)$, which gives us the solution (after insertion in the defintion of $H(Z)$).
Example : $\mathcal{X} = \{1, 2, 3\}$ and $\mathcal{Y} = \{10, 20, 30, 40\}$ for any probability distribution of X and Y, where X and Y are independently distributed.

Optional.

P(X,Y)		$X = -1$	0	1	P(Y)
$Y =$	-2	0	1/3	0	1/3
	1	1/3	0	1/3	2/3
P(X)		1/3	1/3	1/3	1

In this example, one verifies $\langle X\rangle = \langle Y\rangle = \langle XY\rangle = 0$, thus $r = 0$.
$H(X : Y) = H(X) + H(Y) - H(X, Y) = 0.918$ bit.