

## Solutions to Exercise Sheet 1

### Exercise 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) 6 bits are needed since  $2^6 = 64$ .
- (d) 3 symbols are needed since  $4^3 = 64$ .
- (e) Define 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:

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

It follows that  $p_i = 2^{\lambda - \frac{1}{\ln 2}}$ , i.e.  $p_i$  is a constant. If the constraint is applied then  $p_i = \frac{1}{m}$ .

### Exercise 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  $X$ ; therefore it is impossible to compress that source.

### Exercise 3. (a) $H(X) = 2$ bits.

- (b) Sequence of questions:  
Did "head" come up on the first flip?  
Did "head" come up on the second flip??  
:  
Did "head" come up on the  $n$ th flip?  
One bit can be associated with the answer to each question. The answers to  $n$  questions are therefore encoded in  $n$  bits. The expected number of "yes/no" questions is given by  $\sum_{n=1}^{\infty} p(n)n = H(X) = 2$ . It is equal to the entropy, which shows that the sequence of questions is optimal.

### Exercise 4.

- (a)  $H(Y) = H(X) = 1.875$  bits, because the function is bijective (i.e. fixing  $Y$  also fixes  $X$ ).
- (b) The function is not bijective, so  $H(Y) < H(X)$  with  $H(X) = 1/2 + \log_2 3 \approx 2.085$  bits and  $H(Y) = 3/2 + 1/2 \log_2 3 - 5/12 \log_2 5 \approx 1,325$  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.

**Exercise 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:

	Y	-1	-2	-3	-4	P(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)	1/4	1/4	1/4	1/4	

How to compute  $H(X)$  and  $H(Y)$ :

$$H(Y) = H(X) = H(1/4, 1/4, 1/4, 1/4) = \log_2 4 = 2 \text{ bits.}$$

We have  $\mathcal{Z} = \{3, 2, 1, 0, -1, -2, -3\}$  with  $P(Z = 0) = 3/4$ ,  $P(Z = 1) = 1/8$  and  $P(Z = -1) = 1/8$ .  
 All other probabilities are zero.

How to compute  $H(Z)$ :

$$H(Z) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{8} = 1.061 \text{ bits.}$$

Note that  $H(X) > H(Z)$  and  $H(Y) > H(Z)$ .

- (c) We require that  $X$  and  $Y$  are independent and all  $z_{i,j} = x_i + y_j$  are distinct for all pairs  $(i, j)$ . If these conditions are satisfied then  $p_z(i, j) = p_x(i)p_y(j)$ , which gives us the solution (after substituting it in the definition 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.

**Exercise 6. Optional**

	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	

In this example, because  $\langle X \rangle = \langle Y \rangle = \langle XY \rangle = 0$ , which makes  $r = 0$ .  
 $H(X : Y) = H(X) + H(Y) - H(X, Y) = 0.918 \text{ bit.}$