3.1

Given *p(x,y)* with *p(0,0)=p(0,1)=p(1,1)=1/3* y *p(1,0)=0,* find

a) H(X) and H(Y)

b) H(X|Y) and H(Y|X)

c) H(X,Y)

d) I(X;Y) and I(Y;X)

a) $p(x) = \sum_y p(x,y)$ => p(X=0)=p(Y=1)=2/3, p(X=1)=p(Y=0)=1/3

$H(X) = -\sum_x p(x) \log p(x) = 2/3 \log_2 3/2 + 1/3 \log_2 3 = 0.918$ bits = *H(Y)*

b) $H(X|Y) = \sum_y p(y) H(X|Y=y) = \frac{1}{3}\left(1\frac{1}{3}\log 1\right) + \frac{2}{3}\left(\frac{1}{2}\log 2 + \frac{1}{2}\log 2\right)$ =0.666 bits = *H(X|Y)*

c) $H(X,Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} = 3\frac{1}{3}\log 3 = \log 3 = 1.585$ bits

d) *I(X;Y)=I(Y;X)=H(Y)-H(Y|X)* = 0.918 – 0.66 = 0.25 bits

3.2.

For each of the sentences below, reason if it possible that $H(Y|X)=0$ or that $H(Y)=0$

i. The nucleotides composing the genome of an organism are being transmitted through a digital communication system. The actual data transmitted (X) are four numbers corresponding to the 4 possible nucleotides (A=0, C=1, G=2, T=3) and due to errors in the transmission the data received is *Y*=X-3.

ii. The previous signal X is transmitted but now the signal received is $Y = EX$ where E is a random variable that can take the values 1 or -1 with 1/2 probability.

iii. The temperature in a laboratory (X) is transmitted wirelessly. The signal received is $Y = (X - 27^{o})^2$. Assume that the laboratory can have a malfunction in the air conditioning or heating systems, so that it can be very cold or very hot.

i. Knowing the value of X enables us to obtain the value of Y, thus, there is no surprise and H(Y|X)=0

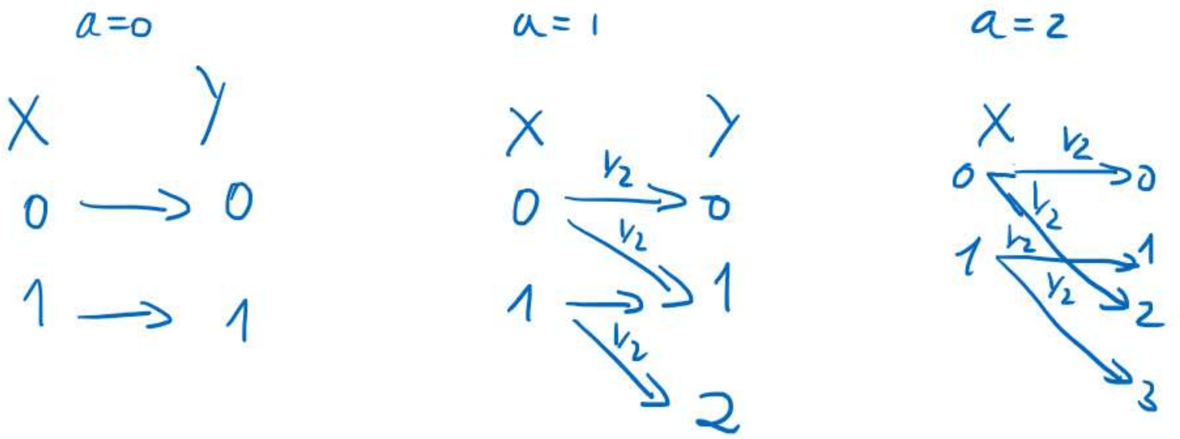
On the other hand, the entropy of X (H(X)) is clearly not zero, since we can assume that each symbol (the nucleotides) has a probability different to zero. Since X=0 always produces Y=-3, X=1 → Y=-2, X=2 → Y=-1, and X=3 → Y=0, the probabilities of the 4 symbol of Y are the same as the probabilities of X, thus H(Y)=H(X) NOT ZERO

iii. If we know the value of X there is no surprise in Y, so H(Y|X)=0. The comment about the fluctuations of the temperature implies that $H(X) \neq 0$. The only way that H(Y)=0 is that it has only one possible value, and this is not the case. So $H(Y) \neq 0$.

3.3.

Given a channel with additive noise. When the symbols {0,1} from an information source X are transmitted, the symbols received are Y=X+Z, where Z is a binary random variable with equally probable values {0, a}, a is an integer number.

Find the capacity of the channels for *a=0, a=1 and a=*2.



a=0

Y=X => C=1 bit

a=1

X=0 → Y=0 or Y=1

X=1 → Y=1 or Y=2

This is the erasure channel. After computing the capacity we get that C=0.5 bits

a=2

$a=1 \qquad C = \max_{P(x)} I(Y;X)$

$$I(Y;X) = \sum_{x,y} p(x,y) \lg \frac{P(x,y)}{P(x)P(y)}$$

$P(X=0)=q \qquad P(X=1)=1-q$

$P(Y=0|X=0)=P(Y=1|X=0)=P(Y=1|X=1)=P(Y=2|X=1)=P(Y=2|X=2)=\frac{1}{2} \rightarrow P(0|2)=P(2|0)=0$

$P(Y=0)=\sum_x P(x,0)=\sum_x \underbrace{P(Y=0|x)}_{1/2}P(x)=\frac{1}{2}q$

$P(Y=1)=\sum_x p(x,1)=\sum_x P(Y=1|x)P(x)=\frac{1}{2}q+\frac{1}{2}(1-q)=\frac{1}{2}$

$P(Y=2)=\frac{1}{2}(1-q)$

$\sum_y P(y)=1$

$P(0,0)=P(y=0|x=0)P(x=0)=\frac{q}{2}=P(0,1)$

$P(1,1)=P(y=1|x=1)P(x=1)=\frac{1-q}{2}=P(1,2)$

$P(0,2)=P(1,0)=0$

$\sum_{x,y} P(x,y)=1$

$$I(Y;X)=\underbrace{\frac{q}{2}\lg\frac{q/2}{q\,q/2}}_{(0,0)}+\underbrace{\frac{q}{2}\lg\frac{\overbrace{q/2}^{1}}{q\,\frac{1}{2}}}_{(0,1)}+\underbrace{\frac{1-q}{2}\lg\frac{(1-q)/2}{(1-q)\frac{1}{2}}}_{(1,1)}+\underbrace{\frac{1-q}{2}\lg\frac{\overbrace{(1-q)/2}^{1}}{(1-q)(1-q)/2}}_{(1,2)}$$

$$=\frac{q}{2}\lg\frac{1}{q}+\frac{1-q}{2}\lg\frac{1}{(1-q)}=\frac{1}{2}\left[q\lg\frac{1}{q}+(1-q)\lg\frac{1}{1-q}\right]=\frac{H(q)}{2}$$

$C=\max_q \frac{H(q)}{2}=\frac{1}{2}$ bits

3.4

Find the capacity in bits of an error-free channel used to transmit 30 symbols.

In an error-free chanel H(X)=H(Y), so I(Y;X)=H(X) and, eventually, C=max I(Y;X)= max H(X)=log #X
So C=log2 30 = 4.9 bits

3.5

Given a discrete r.v. $X \in \{x_1, x_2, x_3, x_4, x_5, x_6\}$ with the following probabilities:

P(X=x1)=0.04

P(X=x2)=0.3

P(X=x3)=0.1

P(X=x4)=0.1

X	P(x)
x_1	0.04
x_2	0.3
x_3	0.1
x_4	0.1
x_5	0.06
x_6	0.4

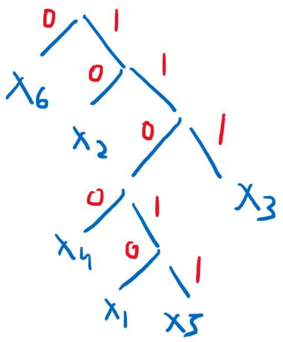
X	P(x)
x_6	0.4
x_2	0.3
x_3	0.1
x_4	0.1
x_1	0.04] 0.1
x_5	0.06

X	P(x)
x_6	0.4
x_2	0.3
x_3	0.1
x_4	0.1
(x_1,x_5)	0.1

X	P(x)
x_6	0.4
x_2	0.3
$x_4(x_1,x_5)$	0.2
x_3	0.1

X	P(x)
x_6	0.4
x_2	0.3
$[x_4(x_1,x_5)]x_3$	

$x_6\ (x_2([x_4\ (x_1,x_5)]\ x_3))$



X	P(x)	C(x)	l_i
x_6	0.4	0	1
x_2	0.3	10	2
x_3	0.1	111	3
x_4	0.1	1100	4
x_1	0.04	11010	5
x_5	0.06	11011	5

$$H(x) = -\sum_x p(x) \lg(x) \simeq 2.14 \text{ bits}$$

$$L = \sum_i l_i \cdot P(x_i) \simeq 2.2 \text{ bits}$$

$$\underset{2.14}{\underbrace{H(x)}} \le \underset{2.2}{\underbrace{L}} \le \underset{2.58}{\underbrace{\lg 6}}$$

UNIT 3 Exercises: Information Theory

3.6

Given a discrete r.v. $Y \in \{Y_1, Y_2\}$ with the following probabilities:
P(Y=y1)=0.7
P(Y=y2)=0.3

1) Compute H(Y)
2) Find a binary Huffman code for Y
3) Compute the average length of the resulting code for Y
4) Apply questions 1, 2 and 3 to a new r.v. Z that includes all possible couples of symbols from Y
5) Using the Huffman code for Z, reason if the average length per Y's symbol is better in contrast with question 3

1, 2, 3

Y	P(Y)	C(Y)	L(C(Y))
Y_1	0.7	0	1
Y_2	0.3	1	1

$H(Y) = -0.7 \log_2 0.7 - 0.3 \log_2 0.3 = 0.88$ bits

$L = E[L(C(Y))] = 0.7 \cdot 1 + 0.3 \cdot 1 = 1$ bit

$\log_2 \#Y = \log_2 2 = 1$ bit

$H(Y) \leq L \leq \log_2 \#Y$

$0.88 \quad 1 \quad 1$

4,5

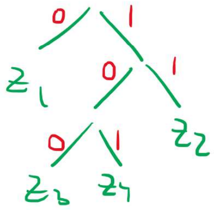
Z	$P(Z)$
$Z_1 = Y_1Y_1$	$0.7^2 = 0.49$
$Z_2 = Y_1Y_2$	$0.7 \cdot 0.3 = 0.21$
$Z_3 = Y_2Y_1$	0.21
$Z_4 = Y_2Y_2$	0.09

$0.21 + 0.09 = 0.3$

Z	$P(Z)$
Z_1	0.49
Z_3Z_4	0.3
Z_2	0.21

$0.3 + 0.21 = 0.51$

$Z_1 \left[(Z_3 Z_4) Z_2 \right]$



Z	$P(Z)$	$C(Z)$	l_i
Z_1	0.49	0	1
Z_2	0.21	11	2
Z_3	0.21	100	3
Z_4	0.09	101	3

$H(Z) = 1.76$ bits

$L_Z = 1.81$ bits $\rightarrow \frac{L_Z}{2} \simeq 0.9 < L_Y$

$\log_2 4 = 2$ bits

$\uparrow$

1

The average length of Z must be divided by two in order to compare it with the average length of Y, since each symbol from Z involves two symbols from Y. Lz/2 is smaller than LY, so the Huffman code for Z is more efficient than the one for Y.
Also, it is important to stress that Z is related to Y. We are using Z to encode Y in a more efficient way. Z and Y are not independent random variables.