

### Exam 2

**Problem No. 1: Entropy Coding**

The alphabet  $A = \{0, 1, 2, 3, 4\}$  has a pdf  $P(A) = \{0.4, 0.3, 0.2, 0.05, 0.05\}$ .

(a) Compute the average codeword length for an optimal code.

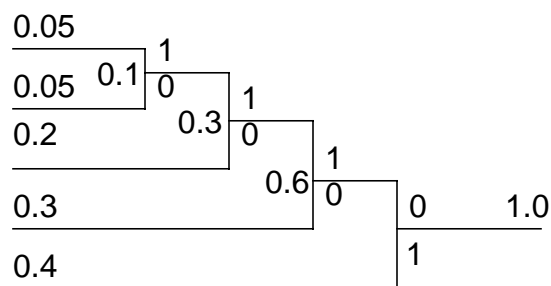
Huffman code is the optimal code and is computed from the probabilities of occurrence of the symbols.

Symbol	P(A)	codeword	length
0	0.4	1	1
1	0.3	00	2
2	0.2	010	3
3	0.05	0110	4
4	0.05	0111	4

$$\text{Average Codeword Length} = 0.4 \times 1 + 0.3 \times 2 + 0.2 \times 3 + 0.05 \times 4 + 0.05 \times 4$$

$$= 2 \text{ bits}$$

Construction of the Huffman Code:



(b) Is there a non-prefix code that has a lower expected code word length than the most optimal prefix code? If so, give an example.

Yes, there can be a non-prefix code which has codeword length less than the most optimal prefix code. As the code need not be uniquely decodable there is no restriction on the assignment of the codewords. Consider the following non-prefix code.



### Example of a non-prefix code

Symbol	P(A)	codeword	length
0	0.4	0	1
1	0.3	1	1
2	0.2	00	2
3	0.05	01	2
4	0.05	10	2

$$\begin{aligned} \text{Average codeword length} &= 0.4 \times 1 + 0.3 \times 1 + 0.2 \times 2 + 0.05 \times 2 + 0.05 \times 2 \\ &= 1.3 \text{ bits} \end{aligned}$$

So, it is demonstrated that a non-prefix code with codeword length less than that of an optimal code exists.

(c) Suppose sending a codeword through this channel costs an amount of money non linearly related to the codeword length:  $Cost = (2)^{l_i}$ . For example the cost of a codeword of length 2 is \$4, and the cost of a codeword of length 3 is \$8. Design a code that delivers a good compression rate, it minimizes the overall cost.

A good compression ratio can be achieved when the bit rate is minimized  
Consider an optimal code i.e, huffman code.

From the codeword length of each symbol, the cost can be computed as:

$$\begin{aligned} Cost &= 2^1 + 2^2 + 2^3 + 2^4 + 2^4 \\ &= 2 + 4 + 8 + 16 + 16 \\ &= \$46 \end{aligned}$$

Consider a non-prefix code: for example take the non-prefix code shown in (b)

From the codeword length of each symbol, the cost for the non-prefix code is

$$Cost = 2^1 + 2^1 + 2^2 + 2^2 + 2^2$$

$$= \$16.$$

Thus the non-prefix code minimizes the cost and also delivers good compression ratio.

Problem No. 2: Binary Symmetric Channels (BSC) and Capacity. For the BSC with transition probabilities  $P(0|1) = P(1|0) = P$ :

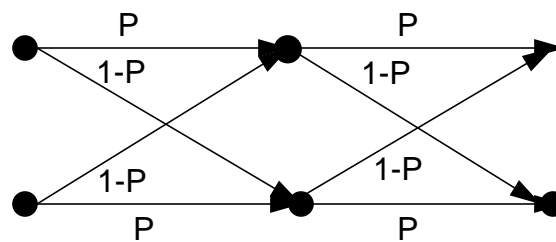
(a) Find the value of  $P$  that minimizes the capacity of the channel. Explain

The mutual information for the Binary Symmetric Channel is given by  $1 - H(p)$ . For the capacity to be minimum,  $H(P)$  should be maximum. The maximum value of  $H(P)$  is 1. We know that  $H(P)$  is 1 when  $p = 0.5$

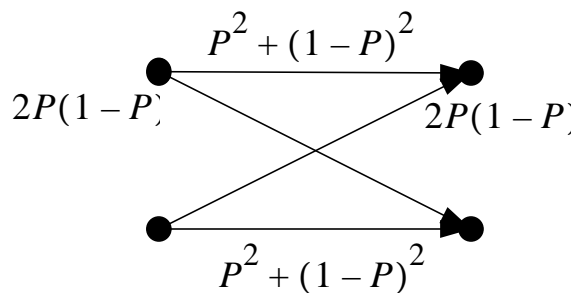
Hence the value of  $P$  that minimizes the capacity of a binary symmetric channel is 0.5

(b) Compute the capacity of the cascade of two channels.

Two Binary Symmetric channels connected in cascade is given below:



The equivalent as a single symmetric channel is



The capacity of the cascade of two channels is

$$1 - H(P^2 + (1-P)^2)$$



(c) Explain to what value the capacity of  $n$  identical BSC channels converges as  $n$  becomes large

We have seen from the capacity of the two cascaded channels that  $H(P)$  increases when compared to the single channel. Similarly as  $n$  increases,  $H(P)$  increases still further and capacity decreases. Hence as  $n$  is sufficiently large the capacity converges to zero.

### Problem No. 3: Differential Entropy

(a) A continuous random variable has a pdf:  $f(x) = e^{-x}$  where  $x \geq 0$ . Compute the entropy.

$$H(x) = -\int_0^{\infty} f(x) \log f(x) dx$$

$$= -\int_0^{\infty} e^{-x} (\log e^{-x}) dx$$

$$= -\int_0^{\infty} \frac{e^{-x} \ln e^{-x}}{\ln 2} dx$$

$$= -\frac{1}{\ln 2} \int_0^{\infty} e^{-x} (-x) dx$$

$$= -\frac{1}{\ln 2} \int_0^{\infty} x e^{-x} dx$$

using integration by parts,

$$\begin{aligned}
 H(x) &= \frac{1}{\ln 2} [-e^{-x}(1+x)]_0^{\infty} \\
 &= \frac{1}{\ln 2} \text{nats}
 \end{aligned}$$

(b) A Continuous random variable has a pdf:  $g(x) = 2e^{-2x}$  where  $x \geq 0$ . Compute the entropy and compare to (a). Explain any differences.

$$\begin{aligned}
 H(x) &= -\int_0^{\infty} f(x) \log f(x) dx \\
 &= -\int_0^{\infty} 2e^{-2x} (\log 2e^{-2x}) dx \\
 &= -\int_0^{\infty} \frac{2e^{-2x} \ln 2e^{-2x}}{\ln 2} dx \\
 &= -\frac{2}{\ln 2} \int_0^{\infty} e^{-2x} (\ln 2 - 2x) dx \\
 &= -\frac{2}{\ln 2} \int_0^{\infty} (e^{-2x} \ln 2 - 2xe^{-2x}) dx
 \end{aligned}$$

Integrating by parts, we get

$$H(x) = -1 + \frac{1}{\ln 2} \text{nats}$$

The difference in the values from (a) is due to the scaling theorem. According to the scaling theorem  $H(ax) = H(x) + \log|a|$ . hence the difference from the answer of the first part.

(c) A continuous random variable has a pdf:  $H(x) = e^{-x+3}$  where  $x \geq 3$ . Compute entropy. Explain.

Translation theorem can be used to compute the entropy. Translation theorem states  $H(x+c) = H(x)$ . Since  $e^{-x+3}$  is the translation of  $e^{-x}$ , the entropy is the same. Hence,

$$H(e^{-x+3}) = H(x) = \frac{1}{\ln 2} \text{nats}$$

(d) Compute the relative entropy between  $f(x)$  (in part a) and  $g(x)$  (in part b):  $D(f||g)$ .

$$\begin{aligned} D(f||g) &= \int_0^{\infty} f \log \frac{f}{g} dx \\ &= \int_0^{\infty} e^{-x} \log \frac{e^{-x}}{2e^{-2x}} dx \\ &= \int_0^{\infty} e^{-x} \left( \log \frac{e^{-x+2x}}{2} \right) dx \\ &= \int_0^{\infty} e^{-x} (\log e^x - \log 2) dx \end{aligned}$$

$$= \frac{1}{\ln 2} \int_0^{\infty} e^{-x} x - \int_0^{\infty} e^{-x} (\log 2) dx$$

Integrating by parts, we get

$$D(f||g) = -1 + \frac{1}{\ln 2}$$