EE8993 Information Theory

Exam no. 2

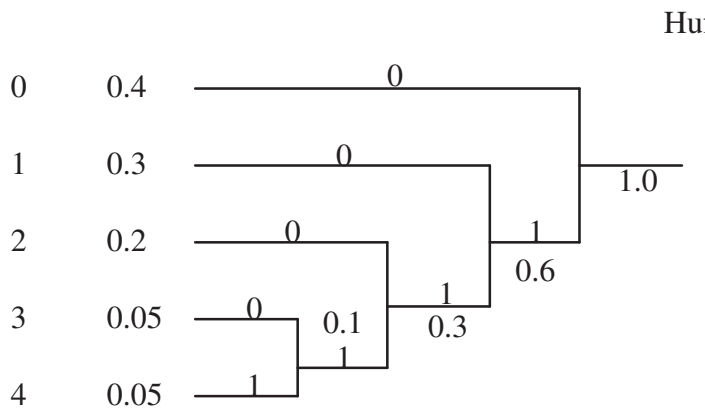
Vinod K. Valsalam

MS Computational Engineering

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 optimal. The expected length of this code is,

$$L = E[l(x)]$$
$$= 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * (0.5 + 0.5)$$
$$= 2 \text{ bits}$$

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

There might be a non–prefix code that has lower expected length than the most optimal prefix code. The most optimal prefix code has to satisfy

$$H(X) \leq L \leq H(X) + 1$$

where L is the expected length of the optimal code. This inequality says that the expected length is greater than or equal to the entropy. If it is greater, we can find a code that has lower expected length than the optimal prefix code, because entropy is the minimum expected number of bits required to encode the source. If L = H(X), then there does not exist a non–prefix code that has lower expected length. L = H(X) is attained if p(x) is dyadic.

For example, the entropy of A for the distribution given in part (a) is

$$H(X) = -(0.4log0.4 + 0.3log0.3 + 0.2log0.2 + 0.05log0.05 + 0.05log0.05)$$
$$= 1.9464 \text{ bits}$$

Since H(X) is less than the expected length of the optimal prefix code that was developed in part (a), it is possible to construct a code that has lower expected length.

In fact, if we relax the condition of unique decodability of the code, we can even construct a code which has expected length less than the entropy. Consider the codeword assignment {(0, 0), (1, 1), (2, 10), (3, 11), (4, 01)}. The expected length of this code is

$$E[l(x)] = 0.4 * 1 + 0.3 * 1 + 0.2 * 2 + 0.05 * 2 + 0.05 * 2 = 1.3 \text{ bits}$$

which is less than the expected length of the optimal code developed in the previous section and the entropy.

(c) Suppose sending a codeword through this channel costs an amount of money nonlinearly related to the codeword length : Cost = 2^{li}. 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, yet minimizes the overall cost.

We have to minimize the cost function, C = 2^{li} subject to the constraint $\sum 2^{-l_i} \leq 1$ (Kraft inequality).

Using Lagrange multiplier method, the function that we have to minimize is

$$J = \sum p_i 2^{l_i} + \lambda\left(\sum 2^{-l_i}\right)$$

$$\frac{dJ}{dl_i} = p_i \ln 2\, 2^{l_i} - \lambda \ln 2\, 2^{-l_i} = 0$$

$$p_i 2^{l_i} - \lambda 2^{-l_i} = 0$$

$$p_i 2^{2l_i} = \lambda$$

$$l_i = \frac{1}{2}\log_2 \frac{\lambda}{p_i} = k_1 - k_2 \log_2 p_i$$

where k_1 and k_2 are constants.

Thus it is seen that l_i is related linearly to log p_i. An optimal code also has the same property. (Optimal code lengths are given by $l_i^* = -\log p_i$) Therefore, the optimal prefix code developed in section (a) will also minimize the overall cost.

Problem no. 2 : Binary Symmetric Channels (BSC) and Capacity

For the BSC with transition probabilities p(0 | 1) = p(1 | 0) = p :

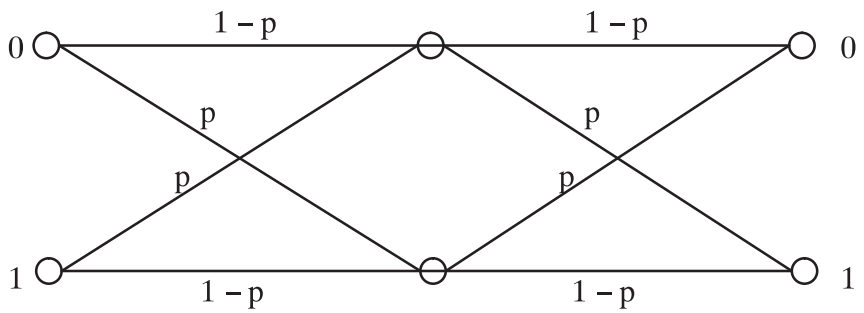
(a) Find the value of p that minimizes the capacity of the channel. Explain.

Capacity for a BSC is

$$C = 1 - H(p)$$

Capacity is minimum when H(p) is maximum. The maximum value of H(p) is 1, which happens when p = 0.5. The minimum value of capacity is therefore 0 which is realized for p = 0.5.

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



The cascade of the two BSC's can be treated as a single BSC for the purpose of analysis with transition probabilities

$p(0 \mid 1) = p(1 \mid 0) = 2p(1 - p) = p_2$

The capacity of the cascade is

$C = 1 - H(p_2) = 1 - H(2p(1 - p))$

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

As n increases, the entropy of transition also increases and attains a value of 1 bit when n becomes infinity. Following the approach taken in the previous section, the n identical BSC's can be treated as a single BSC with transition probabilities $p(0 \mid 1) = p(1 \mid 0) = p_n$. Capacity of the n channels can be written as

$$C = 1 - H(p_n)$$

Since $H(p_n)$ approaches 1 when n becomes large, capacity converges to 0.

Problem no. 3 : Differential Entropy

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

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

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

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

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

$$= \frac{1}{\ln 2}[xe^{-x} - e^{-x}]_0^\infty$$

$$= \frac{1}{\ln 2}\text{ bits}$$

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

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

$$= -\int_0^\infty 2e^{-2x}(log2 + \log e^{-2x})dx$$

$$= -\int_0^\infty 2e^{-2x} * 1dx - \int_0^\infty 2e^{-2x}\log e^{-2x}dx$$

Substitute $y = 2x$, which means that $dy = 2dx$

$$= -2\left[\frac{e^{-2x}}{-2}\right]_0^\infty - \int_0^\infty e^{-y}\log e^{-y}dx$$

$$= -1 + \frac{1}{\ln 2}\text{ bits}$$

g(x) can be written as

$$g(x) = 2f(2x)$$

Entropy when the random variable is scaled by 2 is given by

$$h(2x) = \log 2 + h(x)$$

Comparing the entropy of g(x) with that of f(x) we find that they differ by a –1, which results from the scaling of f(x) to get g(x).

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

With respect to the function, f(x) given in part (a), g(x) can be written as

$$g(x) = f(x - 3)$$

Adding a constant does not change the entropy :

$$h(x - 3) = h(x)$$

Therefore, the entropy of g(x) is

$$h(x) = \frac{1}{\ln 2} \text{ bits}$$

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

$$D(f \| g) = \int_0^\infty f(x) \log \frac{f(x)}{g(x)} dx$$

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

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

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

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

$$= -1 + \frac{1}{\ln 2} \text{ bits}$$