

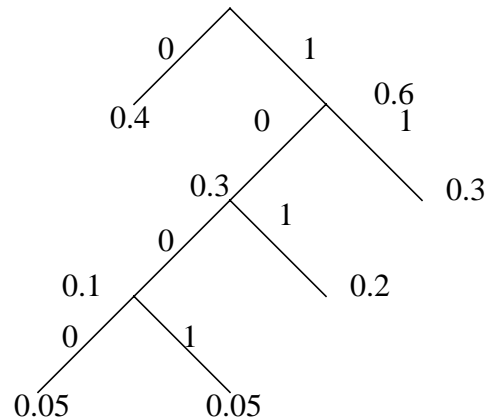
## SOLUTION TO EXAM NO. 2

### Problem No. 1: Entropy Encoding

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.

We proved in class that Huffman code is an optimal code.



So we have

**Table 1: Huffman Code for Alphabet A with pdf  $p(A)$**

A	$p(A)$	Code
0	0.4	0
1	0.3	11
2	0.2	101
3	0.05	1000
4	0.05	1001

$$\begin{aligned} E[l(x)] &= \sum_{x=1}^n p(x)l(x) \\ &= 0.4 \cdot 1 + 0.3 \cdot 2 + 0.2 \cdot 3 + 0.05 \cdot 4 + 0.05 \cdot 4 = 2 \text{ bits} \end{aligned}$$

(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.

Yes, try this:

**Table 2: A Non-uniquely Decodable Code for Alphabet A with pdf p(A)**

A	p(A)	Code
0	0.4	0
1	0.3	1
2	0.2	01
3	0.05	10
4	0.05	11

$$E[l(x)] = \sum_{x=1}^n p(x)l(x)$$

$$= 0.4 \cdot 1 + 0.3 \cdot 1 + 0.2 \cdot 2 + 0.05 \cdot 2 + 0.05 \cdot 2 = 1.3bits$$

(c) Suppose sending a codeword through this channel costs an amount of money nonlinearly 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, yet minimize the overall cost.

To get better compression ratio, we need lower bit rate. To minimize cost, we also need average length to be low. From (b), we can see that Huffman code is not minimum.

If we were to use the Huffman code from (a), we have this:

$$E[c(x)] = \sum_{x=1}^n p(x)l(x)c(x)$$

$$= 0.4 \cdot 1 \cdot 2^1 + 0.3 \cdot 2 \cdot 2^2 + 0.2 \cdot 3 \cdot 2^3 + 0.05 \cdot 4 \cdot 2^4 + 0.05 \cdot 4 \cdot 2^4$$

$$= 14.4dollars$$

If we were to use the non-uniquely decodable code from (b), we have this:

$$\begin{aligned}
 E[c(x)] &= \sum_{x=1}^n p(x)l(x)c(x) \\
 &= 0.4 \cdot 1 \cdot 2^1 + 0.3 \cdot 1 \cdot 2^1 + 0.2 \cdot 2 \cdot 2^2 + 0.05 \cdot 2 \cdot 2^2 + 0.05 \cdot 2 \cdot 2^2 \\
 &= 3.8 \text{dollars}
 \end{aligned}$$

So do you want to spend \$14.4 or \$3.8?

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

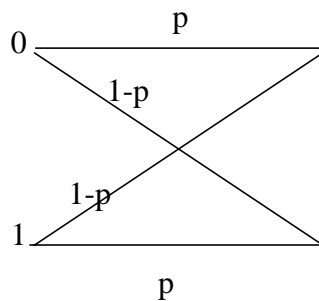
For the BSC with transition probabilities  $p(0/1) = p(1/0) = p$

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

For BSC,  $C = 1 - H(p)$ . We want  $C$  to be minimum so we have to maximize  $H(p)$ .  $H(p)$  is maximum when  $p$  is equally probable (completely random)  $p = 0.5$ .

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

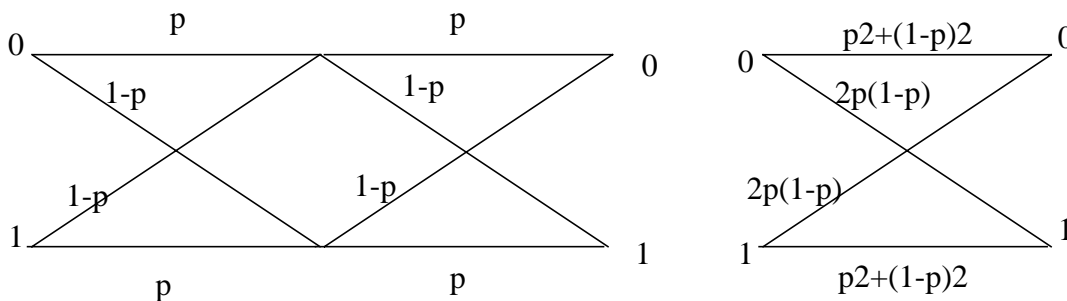
For a non cascading BSC,



the capacity is given by:

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

To find the capacity of a cascade of two of these channels, we reduce the cascade into a non-cascade equivalent.



We have

$$C = 1 - H(p^2 + (1-p)^2)$$

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

From (b) we can see that as we cascade more and more channels, we have more and more chances of errors. For BSC channels, there are two paths to 0 whereas for a one stage cascading of BSC channels, there are four paths to 0. More paths lead to more errors. You can also see this in the equations in (b). The entropy will be greater as more cascading stages are added thus reduces the capacity. The capacity will go to 0 as  $n$  goes to infinity.

Problem No. 3: Differential Entropy

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

$$\begin{aligned} h(X) &= - \int_{x=0}^{\infty} e^{-x} \log e^{-x} dx \\ &= \frac{-1}{\ln 2} \int_{x=0}^{\infty} (-x) e^{-x} dx \end{aligned}$$

Using integration by parts, we have

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

(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.

Let  $y(x) = e^{-2x}$ . Using technique in (a), we find the entropy for  $y(x)$ .

$$\begin{aligned}
 h_y(X) &= - \int_{x=0}^{\infty} e^{-2x} \log e^{-2x} dx \\
 &= \frac{1}{\ln 2} \int_{x=0}^{\infty} (-2x) e^{-2x} dx
 \end{aligned}$$

Using integration by parts, we have

$$\begin{aligned}
 h_y(X) &= \frac{-1}{\ln 2} \left[ x e^{-2x} + \frac{1}{2} e^{-2x} \right]_0^{\infty} \\
 &= \frac{1}{2 \ln 2}
 \end{aligned}$$

From scaling theorem,  $h(aX) = h(X) + \log|a|$

we can find the entropy of  $g(x)$  by scaling the entropy for  $y$  which gives

$$\begin{aligned}
 h_g(X) &= \frac{1}{2 \ln 2} + \log 2 \\
 &= \frac{1}{2 \ln 2} + 1
 \end{aligned}$$

We notice that  $g(x)$  increases entropy.

(c) A continuous random variable has a pdf:  $g(x) = 2e^{-2x}$  where  $x \geq 3$ . Compute the entropy. Explain.

$$h(X) = - \int_{x=3}^{\infty} e^{-x+3} \log e^{-x+3} dx$$

We use the translation theorem proved in class

$$h(X + c) = h(X)$$

Translation does not change the differential entropy. Thus,  $h(X) = h(X)$  of part (a) which is  $1/\ln 2$ .

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

$$\begin{aligned} D(f \parallel g) &= \int_{(x=0)}^{\infty} e^{-x} \log \frac{e^{-x}}{2e^{-2x}} dx \\ &= \frac{1}{\ln 2} \int_{(x=0)}^{\infty} e^{-x} \left( x + \ln \frac{1}{2} \right) dx \\ &= \left( \frac{1}{\ln 2} \int_{(x=0)}^{\infty} x e^{-x} dx + \frac{1}{\ln 2} \int_{(x=0)}^{\infty} \ln \frac{1}{2} e^{-x} dx \right) \\ &= \left[ \frac{-1}{\ln 2} (x e^{-x} + e^{-x}) + \frac{-1}{\ln 2} \ln \frac{1}{2} e^{-x} \right]_0^{\infty} \\ &= \frac{1}{\ln 2} + 1 \end{aligned}$$