# Tutorial — $16^{th}$ October

## Section - 1

choice of $\delta$

$\varepsilon \rightarrow \varepsilon'$

1. $X \sim \mathrm{Bern}(\delta)$

$L = 1$

$H(X) = H_2(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$

$\Rightarrow H(X) \longrightarrow 0$ (say, $\varepsilon'$)

then $L(X) > H(X) + 1 - \varepsilon$

2. a. (a) b. (b, c) ] (c) 2 at a time

(a) Applying the Huffman algorithm gives us the following table

| Code | Symbol | Probability | | | |
|---|---|---|---|---|---|
| 0 | 1 | 1/3 | 1/3 | 2/3 | 1 |
| 11 | 2 | 1/3 | 1/3 | 1/3 | |
| 101 | 3 | 1/4 | 1/3 | | |
| 100 | 4 | 1/12 | | | |

which gives codeword lengths of 1,2,3,3 for the different codewords.

(b) Both set of lengths 1,2,3,3 and 2,2,2,2 satisfy the Kraft inequality, and they both achieve the same expected length (2 bits) for the above distribution. Therefore they are both optimal.

(c) The symbol with probability $1/4$ has an Huffman code of length 3, which is greater than $\lceil \log \frac{1}{p} \rceil$. Thus the Huffman code for a particular symbol may be longer than the Shannon code for that symbol. But on the average, the Huffman code cannot be longer than the Shannon code.

3.

Since the distribution is uniform the Huffman tree will consist of word lengths of $\lceil \log(100) \rceil = 7$ and $\lfloor \log(100) \rfloor = 6$. There are 64 nodes of depth 6, of which (64-$k$) will be leaf nodes; and there are $k$ nodes of depth 6 which will form $2k$ leaf nodes of depth 7. Since the total number of leaf nodes is 100, we have

$$(64 - k) + 2k = 100 \Rightarrow k = 36.$$

So there are 64 - 36 = 28 codewords of word length 6, and $2 \times 36 = 72$ codewords of word length 7.

4.

(a) $H(p) = \frac{1}{2} \log 2 + \frac{1}{4} \log 4 + \frac{1}{8} \log 8 + \frac{1}{16} \log 16 + \frac{1}{16} \log 16 = 1.875$ bits.
$H(q) = \frac{1}{2} \log 2 + \frac{1}{8} \log 8 + \frac{1}{8} \log 8 + \frac{1}{8} \log 8 + \frac{1}{8} \log 8 = 2$ bits.
$D(p||q) = \frac{1}{2} \log \frac{1/2}{1/2} + \frac{1}{4} \log \frac{1/4}{1/8} + \frac{1}{8} \log \frac{1/8}{1/8} + \frac{1}{16} \log \frac{1/16}{1/8} + \frac{1}{16} \log \frac{1/16}{1/8} = 0.125$ bits.
$D(p||q) = \frac{1}{2} \log \frac{1/2}{1/2} + \frac{1}{8} \log \frac{1/8}{1/4} + \frac{1}{8} \log \frac{1/8}{1/8} + \frac{1}{8} \log \frac{1/8}{1/16} + \frac{1}{8} \log \frac{1/8}{1/16} = 0.125$ bits.

(b) The average length of $C_1$ for $p(x)$ is 1.875 bits, which is the entropy of $p$. Thus $C_1$ is an efficient code for $p(x)$. Similarly, the average length of code $C_2$ under $q(x)$ is 2 bits, which is the entropy of $q$. Thus $C_2$ is an efficient code for $q$.

(c) If we use code $C_2$ for $p(x)$, then the average length is $\frac{1}{2} * 1 + \frac{1}{4} * 3 + \frac{1}{8} * 3 + \frac{1}{16} * 3 + \frac{1}{16} * 3 = 2$ bits. It exceeds the entropy by 0.125 bits, which is the same as $D(p||q)$.

(d) Similary, using code $C_1$ for $q$ has an average length of 2.125 bits, which exceeds the entropy of $q$ by 0.125 bits, which is $D(q||p)$.

## Section 2

1a.

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) && (7.2)\\
&= H(Y) - \sum p(x)H(Y|X=x) && (7.3)\\
&= H(Y) - \sum p(x)H(p) && (7.4)\\
&= H(Y) - H(p) && (7.5)\\
&\leq 1 - H(p), && (7.6)
\end{aligned}
$$

where the last inequality follows because $Y$ is a binary random variable. Equality is achieved when the input distribution is uniform. Hence, the information capacity of a binary symmetric channel with parameter $p$ is

$$C = 1 - H(p) \qquad \text{bits.} \tag{7.7}$$

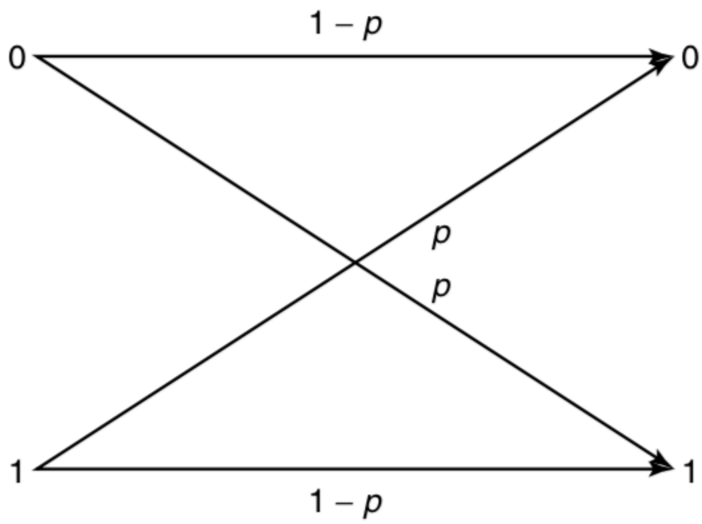


**FIGURE 7.5.** Binary symmetric channel. $C = 1 - H(p)$ bits.

1b.

where $Z$ has some distribution on the integers $\{0, 1, 2, \ldots, c-1\}$, $X$ has the same alphabet as $Z$, and $Z$ is independent of $X$.

In both these cases, we can easily find an explicit expression for the capacity of the channel. Letting $\mathbf{r}$ be a row of the transition matrix, we have

$$I(X;Y) = H(Y) - H(Y|X) \tag{7.19}$$

$$= H(Y) - H(\mathbf{r}) \tag{7.20}$$

$$\leq \log|\mathcal{Y}| - H(\mathbf{r}) \tag{7.21}$$

with equality if the output distribution is uniform. But $p(x) = 1/|\mathcal{X}|$ achieves a uniform distribution on $Y$, as seen from

$$p(y) = \sum_{x \in \mathcal{X}} p(y|x)p(x) = \frac{1}{|\mathcal{X}|}\sum p(y|x) = c\frac{1}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|}, \tag{7.22}$$

where $c$ is the sum of the entries in one column of the probability transition matrix.

Thus, the channel in (7.17) has the capacity

$$C = \max_{p(x)} I(X;Y) = \log 3 - H(0.5, 0.3, 0.2), \tag{7.23}$$

and $C$ is achieved by a uniform distribution on the input.

The transition matrix of the symmetric channel defined above is doubly stochastic. In the computation of the capacity, we used the facts that the rows were permutations of one another and that all the column sums were equal.

Considering these properties, we can define a generalization of the concept of a symmetric channel as follows:

***Definition*** A channel is said to be *symmetric* if the rows of the channel transition matrix $p(y|x)$ are permutations of each other and the columns are permutations of each other. A channel is said to be *weakly symmetric* if every row of the transition matrix $p(\cdot|x)$ is a permutation of every other row and all the column sums $\sum_x p(y|x)$ are equal.

For example, the channel with transition matrix

$$p(y|x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix} \tag{7.24}$$

is weakly symmetric but not symmetric.

is weakly symmetric but not symmetric.

(a) The statistician calculates $\tilde{Y} = g(Y)$. Since $X \to Y \to \tilde{Y}$ forms a Markov chain, we can apply the data processing inequality. Hence for every distribution on $x$,

$$I(X;Y) \geq I(X;\tilde{Y}). \tag{7.1}$$

Let $\tilde{p}(x)$ be the distribution on $x$ that maximizes $I(X;\tilde{Y})$. Then

$$C = \max_{p(x)} I(X;Y) \geq I(X;Y)_{p(x)=\tilde{p}(x)} \geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} = \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}. \tag{7.2}$$

Thus, the statistician is wrong and processing the output does not increase capacity.

(b) We have equality (no decrease in capacity) in the above sequence of inequalities only if we have equality in the data processing inequality, i.e., for the distribution that maximizes $I(X;\tilde{Y})$, we have $X \to \tilde{Y} \to Y$ forming a Markov chain.

4. **Channel capacity.** Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ 1/3, & 1/3, & 1/3 \end{pmatrix}$$

and $X \in \{0, 1, \ldots, 10\}$. Assume that $Z$ is independent of $X$.

(a) Find the capacity.

(b) What is the maximizing $p^*(x)$?

**Solution:** *Channel capacity.*

$$Y = X + Z (\text{mod } 11) \tag{7.15}$$

where

$$Z = \begin{cases} 1 & \text{with probability} 1/3 \\ 2 & \text{with probability} 1/3 \\ 3 & \text{with probability} 1/3 \end{cases} \tag{7.16}$$

In this case,

$$H(Y|X) = H(Z|X) = H(Z) = \log 3, \tag{7.17}$$



independent of the distribution of $X$, and hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) && (7.18) \\ &= \max_{p(x)} H(Y) - H(Y|X) && (7.19) \\ &= \max_{p(x)} H(Y) - \log 3 && (7.20) \\ &= \log 11 - \log 3, && (7.21) \end{aligned}$$

which is attained when $Y$ has a uniform distribution, which occurs (by symmetry) when $X$ has a uniform distribution.

(a) The capacity of the channel is $\log \frac{11}{3}$ bits/transmission.

(b) The capacity is achieved by an uniform distribution on the inputs. $p(X = i) = \frac{1}{11}$ for $i = 0, 1, \ldots, 10$.

(a) Since the channel is symmetric, it is easy to compute its capacity:

$$\begin{aligned} H(Y|X) &= 1 \\ I(X;Y) &= H(Y) - H(Y|X) = H(Y) - 1 . \end{aligned}$$

So mutual information is maximized when $Y$ is uniformly distributed, which occurs when the input $X$ is uniformly distributed. Therefore the capacity in bits is $C = \log_2 5 - 1 = \log_2 2.5 = 1.32$.