Week 5

1 Some comments on engineering quantitites

1.1 Achievability

There exists some scheme using which we can show that the length of compression, rate or any other quantity of interest happens to be equal to be L or that value which is achievable.

1.2 Converse

No scheme exists which can improve upon some value. (can be upper or lower bound depending on situation)

Eg: Suppose in a running race, the fastest speed a human can run is say 10 m/s. Then

- 'Achievable': There exists some person who can run 10 m/s.
- 'Converse': There exists no person who can run more than 10 m/s.
- 'Matching Converse': No human can run at a speed $10 + \epsilon$, for any $\epsilon > 0$.

2 Channel Coding

Say we are inputting $x \in \mathcal{X}$, $(x_1, \cdots, x_n)$, in a channel and we are getting $y \in \mathcal{Y}$, $(y_1, \cdots, y_n)$, with $\mathcal{Y}$ as output alphabet.

If multiple x_i are mapping to a single y, it signifies a noisy channel as we would be unable to decode accurately.

To make this channel one-one (and therefore ensure correct decoding), we omit some sequences (n-length vectors in $\mathcal{X}^n$) from the set of all transmittable sequences.

This subset of transmittable sequences is called as the 'channel code' (or simply code). Denoted generally by $\mathscr{C}$. Note that, $\mathscr{C} \subseteq \mathcal{X}^n$.

Each vector in $\mathscr{C}$ is called a codeword. Number of bits required to represent $|\mathscr{C}|$ codewords

$$= \log_2 |\mathscr{C}| \text{ bits}$$

$$\text{Rate of the code } \mathscr{C} = \frac{\log_2 |\mathscr{C}|}{n} \text{ bits per channel use (bpcu or b/cu)}$$

Intuitively, higher the rate, more the chance of many-one kind of system and higher the chance of error.

2.1 Probabilistically noisy channel

Also called a random channel or random noise.

For $X = x \in \mathcal{X}$, there will be a probability distribution on the output random variable Y. The coditional distribution on Y given $X = x$,

$$P_{Y/X=x} = \{P(Y = y/X = x) : y \in \mathcal{Y}\}$$

These distributions $P_{Y/X}(y/x)$ $\forall x$ completely characterize or describe the random channel.

Now, we need to think about how to calculate $P(error)$. $X_i's$ are given as input to the random channel which are not independent. Then $(Y_1, \cdots, Y_n)$ is given out as output of the random channel, this is then sent into the decoder which then gives out $\hat{X}$, which is an estimate for X.

When $\hat{X} \neq X$, it is called an error event.

$$P(\text{Decoding error}) = P(\hat{X} \neq X)$$

Eg: $\mathcal{X} = \{0, 1\}$, $\mathscr{C} = \{000, 111\}$ instead of all 8 sequences. $P(\text{error})$ decreases but rate of code $= \frac{\log_2 |\mathscr{C}|}{n} = \frac{1}{3}$

Intuitively, it seems like if we want to decrease $P(\text{error})$ we have to increase n and we can expect the error to be close to 0 but this isn't the case.

For any small $\epsilon > 0$, there exists a code $\mathscr{C}$ with $P(\text{error}) \leq \epsilon$ & rate of the code $R(\mathscr{C}) = \max_{P_X}(I(X;Y)) - f(\epsilon)$

Note that:

1. I(X;Y) depends on $P_{Y/X=x}$ $\forall$ $x \in \mathcal{X}$.
2. I(X;Y) depends on distribution of P_X and P_Y.
3. X isn't a natural source upon which we have no control but it is the output of some encoding which encodes the 'raw source'.

   So $P_X(x)$ is generally assumed to be controllable in the mathematical framework of information theory.

The quantity $\max_{P_X}(I(X;Y))$ is called the 'Channel capacity', denoted by C.

3 Channel coding theorem

No matter what we do, we can't get a code with rate $> C$ (channel capacity) and expect a small probability of error.

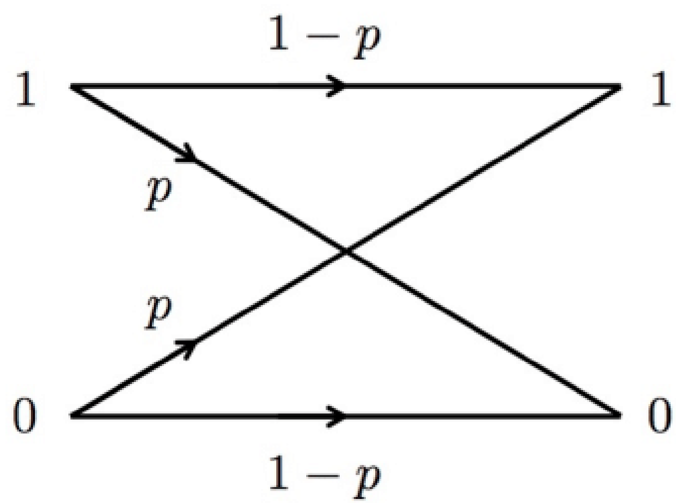
Note: To make the rate very close to C, we have to use a very high value of code length n.

3.1 Binary symmetric channel

$$\mathcal{X} = \{0, 1\} = \mathcal{Y}$$

A binary symmetric channel with crossover probability p, denoted by BSC(p), is a channel with binary input and binary output and probability of error p . That is, if X is the transmitted random variable and Y the received variable, then the channel is characterized by the conditional probabilities:

$$P(Y = 0/X = 0) = 1 - p$$
$$P(Y = 0/X = 1) = p$$
$$P(Y = 1/X = 0) = p$$
$$P(Y = 1/X = 1) = 1 - p$$



We want to find channel capacity of this channel,

$$C = \max_{P_X}(I(X; Y))$$

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y/X) \\ &= H(Y) - \sum_{x \in \{0,1\}} P_X(x) H(Y/X = x) \\ &= H(Y) - \sum_{x \in \{0,1\}} P_X(x) H_2(p) \\ &= H(Y) - H_2(p) \end{aligned}$$

As we know $max(H(Y)) \leq Y$,

$$C_{BSC} = 1 - H_2(p)$$

where $H_2(p)$ is the binary entropy function defined by,

$$H_2(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$$

Hence if P_x is a uniform distribution, the channel capacity can be 1.