Information Theory	**Prof. I-Hsiang Wang**
Lecture Notes: Channel Coding	
November 14, 2015	**Scribe: Chi-Ning Chou**

1 Channel coding theorem

Theorem 1 (channel coding theorem) *Let $C = \max_{p_X} I(X;Y)$, then*

- *(***achievability***) $\forall R \leq C$, there's a set of lossless encoding schemes.*
- *(***converse***) $\forall R$ such that $P_e^{(N)} \to 0$, $R \leq C$.*

1.1 Converse

Now we start the converse proof of Theorem 1. Recall that here we prove the weak converse. That is, $\forall R$ such that $P_e^{(N)} \to 0$, $R \leq C$. We divide the proof into three steps:

1. Since W is uniform, we have

$$NR = K = H(W) = I(W;\hat{W}) + H(W|\hat{W}) \tag{1}$$
$$\leq I(W;Y^N) + (1 + P_e \log|2^K + 1|) \tag{2}$$

   , where the first term is followed by *data processing theorem* and the second term is followed by *Fano's inequality*. Clearly that, the second term will vanish as we divide it by N and let N goes to infinity.

2. Consider the first term in (2), we have

$$\begin{aligned} I(W;Y^N) &= \sum_{k=1}^{N} I(W;Y_k|Y^{k-1}) \\ &\leq \sum_{k=1}^{N} I(W,Y^{k-1};Y_k) \\ (\because X \text{ is generated by } W \text{ and } Y^{k-1}) &= \sum_{k=1}^{N} I(W,Y^{k-1},X_k;Y_k) \\ (\because \text{no feedback}) &= \sum_{k=1}^{N} I(X_k;Y_k) \leq \max_{p_X} I(X;Y) \end{aligned}$$

3. Finally, combine 1. and 2., we have

$$\begin{aligned} R &\leq \frac{1}{N}[\sum_{k=1}^{N} \max_{p_X} I(X;Y) + 1 + P_e(\log|2^K + 1|)] \\ &\leq \max_{p_X} I(X;Y) + \frac{1}{N} + P_e\frac{K+2}{N} \to C \end{aligned}$$

2 Cost constraint

For real application, we might need to address some cost constraint on the channel. For example, the size of alphabet could not be so large, or it consumes more energy to send certain symbols etc. To sum up, we can use a cost function to describe the penalty of the constraint. We denote it as $b(X)$ or $b(X, Y)$, which represent the input cost function and input-output cost function respectively.
However, with these constraint, the channel capacity might decrease. That is, the information we can transmit through the channel cannot be maximized as without the constraint. As a result, we need to consider the capacity under these constraint. Thus, we can write down the cost-constraint channel capacity as

$$C := \sup_{X:\mathbb{E}[X]\leq B} \{R : R \text{ is achievable}\}$$

Here, we adopt an input cost constraint with threshold value B. In the costless channel, the capacity is the supremum of the mutual information among input and output. Here, we might wonder, whether the capacity under constraint is also the supremum mutual information among input and output under the same constraint?
To our desiring, the answer is positive. We can first simply show that every achievable data rate is less than or equal to $\sup_{X:\mathbb{E}[b(X)]\leq B} I(X;Y)$ and we can use typical set to achieve it.