Lecture Notes on October 11

Yiyong Feng

October 14, 2011

1 Background of source code

The following figure Fig. 1 shows the basic encode and decode system.

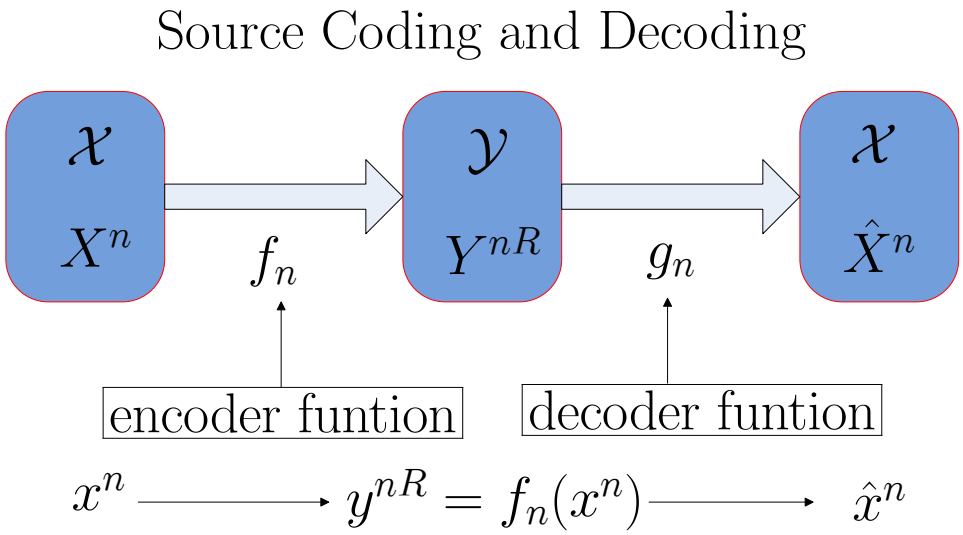

Figure 1: Encode and decode

In which, random variable sequence X^n is the message the transmitter want to send, and each random of the sequence is selected from the set $\mathcal{X}$. Directly transmitting the message may be not possible, we need to define the corresponding codeword of the message that can be transmitted. Assign the messages with different codewords is the encoding. Assume the codebook (the set of all possible codewords) is a random variable Y^{nR} drown from the set $\mathcal{Y}$. The function which maps the message to the codeword is called encoder function, and is denoted as f_n. At the receiver end, the receiver receive the codeword, and he is supposed to recover the message from the codeword. This recovering is done by decoder. The function which maps the codeword to the possible message is called decoder function, denoted as g_n. Of course, there may be some error such that the recovered message is different from the transmitted message.

For example, a particular message x^n is encoded by the function f_n to a codeword y^{nR}, then at the receiver end, y^{nR} is decoded by the function g_n to the possible message $\hat{x}^n$. Hopefully, $\hat{x}^n = x^n$, and whenever this does not hold, an error happens. The pair of functions (f_n, g_n) is called code scheme.

The definition of error probability is

$$P_{X^n}(X^n \neq \hat{X}^n) = \sum_{x^n} P(x^n)\mathbf{1}_{(\hat{x}^n \neq x^n)}. \tag{1}$$

The definition of rate is the average length of the codewords,

$$R = \sum_{x^n} P(x^n) l(y^{nR}), \tag{2}$$

where $y^{nR} = f_n(x^n)$, and the normalized case definition is

$$R = \frac{\sum_{x^n} P(x^n) l(y^{nR})}{n}. \tag{3}$$

2 Converse for source coding theorem

The direct part of source coding says that if the coding rate R is greater than the source entropy $H(X)$, the coding rate tends to $H(x)$ and the error tends to 0. The converse part says that if block code with rate less than the source entropy $H(X)$, the error probability will always exist. The following inequalities will prove the the converse part of source coding.

$$\begin{aligned}
nH(X) &\overset{(a)}{=} H(X^n) \\
&\overset{(b)}{=} H(X^n|\hat{X}^n) + I(X^n; \hat{X}^n) \\
&\overset{(c)}{\leq} H(X^n|\hat{X}^n) + I(X^n; Y^{nR}) \\
&\overset{(d)}{\leq} H(X^n|\hat{X}^n) + H(Y^{nR}) \\
&\overset{(e)}{\leq} H(X^n|\hat{X}^n) + nRH(Y) \\
&\overset{(f)}{\leq} H(X^n|\hat{X}^n) + nR \\
&\overset{(g)}{\leq} 1 + P(X^n \neq \hat{X}^n) n \log |\mathcal{X}| + nR,
\end{aligned}$$

where

- (a) follows from $X_1, \ldots, X_n$ are i.i.d. with entropy $H(X)$;
- (b) follows from identity $H(X) = H(X|Y) + I(X;Y)$;
- (c) follows from Data Processing Inequality (DMI);

- (d) follows from $I(X;Y) \leq \min\{H(X), H(Y)\}$;
- (e) follows from $H(Y^{nR})$ is maximized if all Y_i are i.i.d.;
- (f) follows from binary entropy $H(Y) \leq 1$;
- (g) follows from Fano's inequality.

Thus, if $R < H(X)$, we have

$$P(X^n \neq \hat{X}^n) \geq \frac{nH(x) - nR - 1}{n \log |\mathcal{X}|}, \tag{4}$$

the lower bound when $n \to \infty$ is

$$\lim_{n\to\infty} P(X^n \neq \hat{X}^n) \geq \frac{H(x) - R}{\log |\mathcal{X}|} > 0. \tag{5}$$

The condition is tight when Y is i.i.d. and uniformly distributed and the mapping from $\mathcal{X} \to \mathcal{Y}$ is a one-one mapping. Here, it's important to use fixed-length block coding.

The above converse reflects that when coding rate is less than the source entropy, then there must be some error which can not be eliminated.

3 Achievablility for source coding theorem

To be continued...

4 Converse for channel coding theorem

The channel coding theorem is that any channel rate R_c is achievable for a discrete memoryless channel if and only if $R_c \leq C$, where C is the capacity of the channel,

$$C = \sup_{p(x)} I(X;Y).$$

The converse says that for any $R_c > C$, the error probability must exist and can not be eliminated.

The representation of channel is shown in Fig. 2.

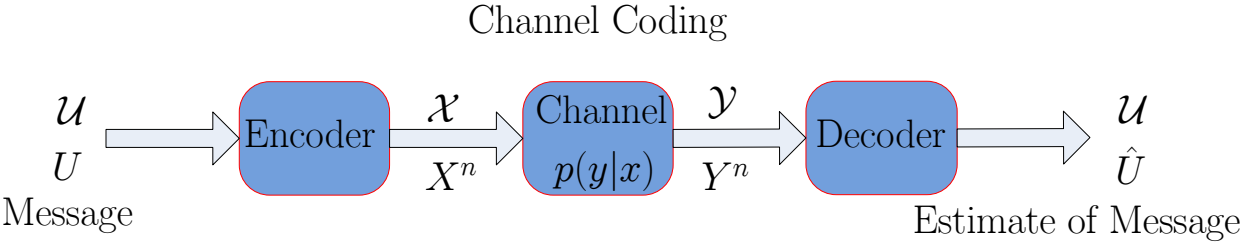

Figure 2: Channel coding

To prove the converse for the channel coding, first we need to prove $I(X^n;Y^n) \leq nC$,

$$\begin{aligned}
I(X^n;Y^n) &\overset{(a)}{=} H(Y^n) - H(Y^n|X^n) \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} H(Y_i) - H(Y^n|X^n) \\
&\overset{(c)}{=} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n) \\
&\overset{(d)}{=} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&\overset{(e)}{=} \sum_{i=1}^{n} I(X_i;Y_i) \\
&\overset{(f)}{\leq} n \sup_{p(x)} I(X;Y) = nC,
\end{aligned} \tag{6}$$

where

- (a) and (e) follow from identity $I(X;Y) = H(X) - H(X|Y)$;
- (b) follows from $H(Y^n)$ is maximized if all Y_i are i.i.d.;
- (c) follows from chain rule;
- (d) follows from the channel is Discrete Memoryless Channel (DMC);
- (f) follows from definition of channel capacity.

Then prove

$$\begin{aligned}
nR &\overset{(a)}{=} H(U) \\
&\overset{(b)}{=} H(U|\hat{U}) + I(U;\hat{U}) \\
&\overset{(c)}{\leq} H(U|\hat{U}) + I(X^n;Y^n) \\
&\overset{(d)}{\leq} 1 + P_e^{(n)} n \log|\mathcal{X}| + nC, \\
&\overset{(a)}{=} 1 + P_e^{(n)} nR + nC
\end{aligned} \tag{7}$$

where

- (a) follows from fixed-length uniformly distribution of X_i;
- (b) follows from identity $H(X) = H(X|Y) + I(X;Y)$;
- (c) follows from DMI;
- (d) follows from Fano's inequality and inequality (6);

From the result (7), it can be derived that

$$P_e^{(n)} \geq 1 - \frac{C}{R} - o(1).$$

Because $R > C$, $P_e^{(n)} > 0$ always holds, then the converse for channel coding is proved.

5 Achievability for channel coding theorem

To be continued...