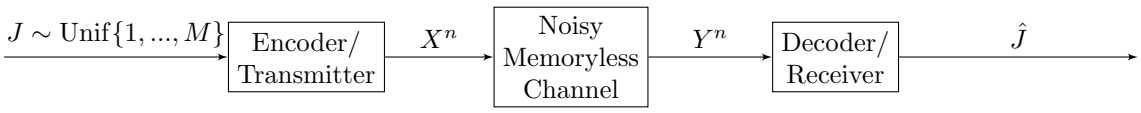
EE376A Information Theory Lecture 12 - 02/17/2015

Lecture 12

Lecturer: Tsachy Weissman *Scribe: Alkan, Choi, Sinha*

1 Communication Setting with a Finite Alphabet



Main Result: $C = C^{(I)} = \max_X I(X;Y)$
Direct Theorem (Lecture 11): If $R < C^{(I)}$, then the rate R is achievable.
Converse Theorem (today's lecture): If $R > C^{(I)}$, then R is not achievable.

1.1 Fano's Inequality

Theorem 1 (Fano's Inequality)**.** *Let X be a discrete random variable and $\hat{X} = \hat{X}(Y)$ be a guess of X based on Y. Let $P_e := P(\hat{X} \neq X)$. Then $H(X|Y) \leq h(P_e) + P_e \log(|\mathcal{X}| - 1)$.*

Proof Let $V = \mathbf{1}_{\{\hat{X} \neq X\}}$.

$$\begin{aligned} H(X|Y) &\leq H(X, V|Y) \quad \text{(Data Processing Inequality)} && (1)\\ &= H(V|Y) + H(X|V, Y) \quad \text{(Chain Rule)} && (2)\\ &\leq H(V) + H(X|V = 0, Y = y)P(V = 0, Y = y) \\ &\quad + H(X|V = 1, Y = y)P(V = 1, Y = y) && (3) \end{aligned}$$

We can simplify terms in (3). First, $H(V) = h(P_e)$, the entropy of a binary random variable with success probability P_e. Furthermore, X is deterministic given $V = 0$ and $Y = y$, so $H(X|V = 0, Y = y) = 0$. Finally, if $V = 1$ and $Y = y$, then $\hat{X}$ is known and X can take up to $|\mathcal{X}| - 1$ values. Thus $H(X|V = 1, Y = y) \leq \log(|\mathcal{X}| - 1)$. Putting these facts together, we arrive at:

$$\begin{aligned} H(X|Y) &\leq h(P_e) + \log(|\mathcal{X}| - 1)P(V = 1) && (4)\\ &= h(P_e) + P_e \log(|\mathcal{X}| - 1) && (5) \end{aligned}$$

□

A weaker version of Fano's Inequality uses the facts that $h(P_e) \leq 1$ and $\log(|\mathcal{X}| - 1) \leq \log(|\mathcal{X}|)$:

$$H(X|Y) \leq 1 + P_e \log(|\mathcal{X}|) \tag{6}$$

or equivalently,

$$P_e \geq \frac{H(X|Y) - 1}{\log(\mathcal{X})} \tag{7}$$

1.2 Converse Theorem

Theorem 2 (Converse Theorem)**.** *If $R > C^{(I)}$, then rate R is not achievable.*

Proof

$$\begin{aligned}
\log M - H(J|Y^n) &= H(J) - H(J|Y^n) = I(J;Y^n) && (8)\\
&= H(Y^n) - H(Y^n|J) && (9)\\
&= \sum_i H(Y_i|Y^{i-1}) - \sum_i H(Y_i|Y^{i-1},J) && (10)\\
&\le \sum_i H(Y_i) - \sum_i H(Y_i|Y^{i-1},J,X^n) \ \text{(conditioning reduces entropy)} && (11)\\
&= \sum_i H(Y_i) - \sum_i H(Y_i|X_i) \ \text{(memorylessness)} && (12)\\
&= \sum_i I(X_i;Y_i) && (13)\\
&\le nC^{(I)} && (14)
\end{aligned}$$

Thus, for schemes with rate $(= \frac{\log M}{n}) \ge R$, we have

$$P_e \ge \frac{H(J|Y^n) - 1}{\log M} \ge \frac{\log M - nC^{(I)} - 1}{\log M} \ge 1 - \frac{C^{(I)}}{R} - \frac{1}{nR} \xrightarrow{n\to\infty} 1 - \frac{C^{(I)}}{R} \qquad (15)$$

If $R > C^{(I)}$, then P_e is bounded from below by a positive constant, so it does not approach 0. Therefore, $R > C^{(I)}$ is not achievable. □

2 Some Notes on the Direct and Converse Theorems

2.1 Communication with Feedback: $X_i(J, Y^{i-1})$

Even if the encoder gets feedback of what has been received on the other side of the channel, one can verify that the proof of converse carries over verbatim; $C = C^{(I)}$ with or without feedback! But, feedback can help improve simplicity and reliability of schemes to achieve the best rate. Here is an example:

Example 3. Communicating Through Erasure Channel

Recall that the capacity of erasure channel is $C = 1 - \alpha$ (bits/channel use).

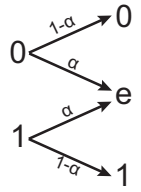


Figure 1: An Erasure Channel

If feedback exists, the transmitter can repeat each information bit until it goes through unerased. On average, one needs $1/(1-\alpha)$ channel uses per information bit. This means that rate achieved in this scheme is $1 - \alpha$ bits/channel use. This simple scheme is completely reliable since the probability of error is equal to zero (every bit will eventually be error-free).

2.2 Practical Schemes

In the proof of direct part, we showed mere existence of C_n (a codebook achieving the rate equivalent to the channel capacity) with a size $|C_n| \geq 2^{nR}$, and small P_e. Even if such C_n is given, encoding and decoding using this codebook for large n are not practical. For practical schemes, see:

1. LDPC Codes: "Low Density Parity Check Codes", Gallager 1963 Thesis [1].
2. Polar Codes: "Channel Polarization", Arikan 2009 [2].
3. Or, take EE388, which will be available in the next academic year.

2.3 P_e vs. $P_{\max}$

In our discussion so far, our notion of reliability has been the (average) probability of error, which is defined as:

$$P_e = P(\hat{J} \neq J) = \frac{1}{M} \sum_{j=1}^{M} P(\hat{J} \neq J | J = j) \tag{16}$$

A more stringent notion of reliability is the maximal probability of error P_{max}, which is defined as:

$$P_{\max} = \max_{1 \leq j \leq M} P(\hat{J} \neq j | J = j) \tag{17}$$

It turns out that our results, i.e., direct and converse theorems are still valid for this more stringent notion of reliability. The converse theorem is clear. If arbitrarily small P_e cannot be achieved, arbitrarily small $P_{\max}$ cannot be achieved either, therefore the converse theorem holds for $P_{\max}$.
We now show that the result of the direct proof holds for vanishing $P_{\max}$. Note that with application of the Markov inequality, we have:

$$|\{1 \leq j \leq M : P(\hat{J} \neq j | J = j) \leq 2P_e\}| \geq \frac{M}{2} \tag{18}$$

Given C_n with $|C_n| = M$ and P_e, there exists C_n' with $|C_n'| = M/2$ and $P_{\max} \leq 2P_e$. By extracting a better half of C_n, one can construct C_n'. The rate of C_n' is:

$$\text{Rate of } C_n' \geq \frac{\log(M/2)}{n} = \frac{\log M}{n} - \frac{1}{n} \tag{19}$$

This implies that if there exists schemes of rate $\geq R$ with $P_e \to 0$, then for any $\epsilon > 0$, there exists schemes of rate $\geq R - \epsilon$ with $P_{\max} \to 0$

References

[1] R. G. Gallager, *Low-Density Parity-Check Codes*, Cambridge, MA: MIT Press, 1963.

[2] E. Arikan, "Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.