Information Theory with Applications, Math6397
Lecture Notes from October 9, 2014

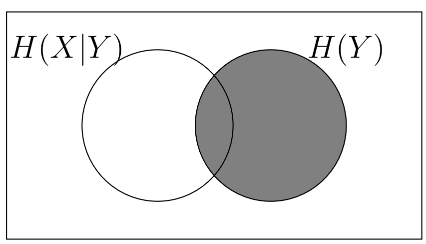
taken by Kedar Grama

Warm Up: Binary erasure channel(BEC)

With $Y = \gamma(X)$ and $\mathbb{P}(\Delta) = \epsilon$

Last time we had observed

$$\begin{aligned}
C &= \max_Y I(X=0;Y) = \max_Y I(X=1;Y) \\
&= \max_{X,Y} I(X;Y) \\
&= \max_{X,Y} (H(Y) - H(Y|X)) \\
&= \max_Y \left(H(Y) - \sum_{i=0}^{1} \mathbb{P}_X(i) H(Y|X=i) \right) \\
&= \max_Y (H(Y) - h(\epsilon))
\end{aligned}$$



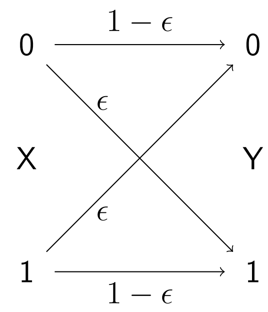
Maximality of $H(Y)$ implies, by the fixed probability $\mathbb{P}_Y(\Delta) = \epsilon$, a uniform distribution on the other two outcomes,

$$\mathbb{P}_Y(0) = \mathbb{P}_Y(1) = \frac{1-\epsilon}{2}\,.$$

We conclude the capacity is:

$$\begin{aligned}
C &= -1\frac{1}{2}(1-\epsilon)\ln\frac{2-1}{2} - \epsilon\ln\epsilon + (1-\epsilon)\ln(1-\epsilon) \\
&= -(1-\epsilon)\ln\frac{1}{2} = (1-\epsilon)\ln 2
\end{aligned}$$

Binary Symmetric Channel



To compute capacity, we compute

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{i=0}^{1} \mathbb{P}_X(i) H(Y|X=i) \\ &= H(Y) - h(\epsilon) \end{aligned}$$

To achieve the max of $I(X;Y)$ we need to maximize $H(Y)$. Entropy is maximized if $\mathbb{P}_Y(0) = \mathbb{P}_X(1) = \frac{1}{2}$, $H(Y) = \ln 2$ Hence, in this case, we transmit at a rate of $R = \ln 2 - h(\epsilon)$.

When $\epsilon = 0$ we transmit as expected at a rate of $\ln 2$, corresponding to one bit per channel use, but the same holds if $\epsilon = 1$, because then the output only has to be inverted to get the input message. Finally, when $\epsilon = 1/2$, the transmission rate is 0 because the channel output is independent of the input.

In order to derive a (weak) converse to the channel coding theorem we use the following lemma.

3.1.16 Lemma (Fano)**.** *Let S, Y be random variables with the finite alphabet $\mathbb{A}$, and*

$$E = \begin{cases} 0, & S = Y \\ 1, & S \neq Y \end{cases}$$

Then,

$$H(S|Y) \leq H(E) + H(S|E,Y) + \mathbb{P}(S \neq Y) \ln |\mathbb{A}| - 1$$

Proof. Express $H(E, S|Y)$ in two ways using additivity

$$H(E,S|Y) = H(S|Y) + \underbrace{H(E|S,Y)}_{=0 \text{ because S and Y determine E}} = H(E|Y) + H(S|E,Y)$$

We estimate:

$$H(S|Y) \leq H(E) + H(S|E,Y)$$

Also, consider

$$\begin{aligned} H(S|E,Y) &= \mathbb{P}_E(0)\underbrace{H(S|Y,E=0)}_{=0} + \underbrace{\mathbb{P}_E}_{\mathbb{P}(S\neq Y)}\ \underbrace{H(S|Y,E=1}_{\leq \ln(|\mathbb{A}|-1)} \\ &\leq \mathbb{P}(S\neq Y)\ln(|\mathbb{A}|-1) \end{aligned}$$

Then, collecting terms we get:

$$H(S|Y) \leq H(E) + \mathbb{P}(S\neq Y)\ln(|\mathbb{A}|-1)$$

□

Next, we state a (weak) converse to the channel coding theorem.

3.1.17 Theorem. *Let γ be a discrete memoryless channel with conditional probabilities $\{\mathbb{W}(b|a)\}, a \in \Delta$ and $\{\mathcal{C}_n, \phi_n, \psi_n\}$ a transmission code sequence with the same size $m_n = |\mathcal{C}_n|$. If,*

$$\lim_{n\to\infty}\inf \frac{1}{n}\ln m_n > C$$

Then, the averaged error prbability P_e can be bounded away from zero for all sufficiently large n.

Proof. Without loss of generality, let $\mathcal{C}_n = \{1,2,\ldots,m_n\}$, $\Phi_n : \mathcal{C}_n \to \mathbb{A}^n$. Assuming equally probable inputs, we have a uniform distribution $\mathbb{Q}$ on $\mathcal{C}_n$ with $H(\mathbb{Q}) = \ln m_n$. Let S be a random variable with values $\mathcal{C}_m$ distributed according to $\mathbb{Q}$. Then, we have $\{S, X=\Phi(S), Y\}$ forming a Markov chain, because

$$\begin{aligned} \mathbb{P}(Y=b) &= \frac{1}{m_n}\sum_{x\in\mathcal{C}_n}\mathbb{W}(b|\Phi_n(x)) \\ &= \sum_{a\in\mathbb{A}^n}\frac{|\Phi^{-1}(a)|}{m_n}\mathbb{W}(b|a) \end{aligned}$$

and X is a deterministic function of S. Now, by the data processing inequality for Markov chains,

$$I(S;Y) \leq I(X;Y)$$

and comparing with a discrete memoryless source as input

$$\begin{aligned} I(X;Y) &\leq \max_{\mathbb{P}_X, Y=\gamma(X)} I(X;Y) \\ &= \max_{\mathbb{P}_X, Y=\gamma(X)} \sum_{j=1}^{n} I(X_j;Y_j) \\ &\leq \max_{\mathbb{P}_X, Y_j=\gamma(X_j)} \sum_{j=1}^{n} \underbrace{I(X_j;Y_j)}_{\leq C} \\ &\leq nC \end{aligned}$$

Now, defining any $\phi_n : \mathbb{B}^n \to \{1, 2, \ldots, m_n\}$ we have for

$$E = \begin{cases} 1, \ \psi_n(Y) \neq S \\ 0, \ \text{else} \end{cases}$$

that

$$\begin{aligned} \ln m_n =& H(S) \\ \overset{additivity}{=}& H(S|Y) + I(S;Y) \\ \overset{Markov}{\neq}& H(S|Y) + I(X;Y) \end{aligned}$$

Next, using Fano's inequality

$$\begin{aligned} \ln m_n &= H(S) \\ &\leq H(E) + \mathbb{P}(E=1)\ln(|\mathcal{C}_n| - 1) + nC \\ &\leq \ln 2 + \underbrace{\mathbb{P}(E=1)}_{P_e}\ln(m_n - 1) + nC \end{aligned}$$

Solving for P_e gives

$$\begin{aligned} P_e &\geq \frac{\ln m_n - nC - \ln 2}{\ln(m_n - 1)} \\ &\geq \frac{\ln m_n - nC - \ln 2}{\ln m_n} \end{aligned}$$

So,

$$P_e \geq 1 - \frac{C}{\frac{1}{n}\ln m_n} - \frac{\ln 2}{\ln m_n}$$

If $\liminf_{n\to\infty} \frac{1}{n}\ln m_n > C$ then there is a $\delta > 0$ and $N \in \mathbb{N}$ such that for all $n > N$, $\frac{1}{n}\ln m_n > C + \delta$

Assuming that N is such that $n > N$ and $\frac{\ln 2}{n} < \frac{\delta}{2}$, then

$$\begin{aligned} P_e &\geq \underbrace{1 - \frac{C}{C+\delta}}_{\frac{\delta}{C+\delta}} - \underbrace{\frac{\ln 2}{n(C+\delta)}}_{\frac{\delta}{2(C+\delta)}} \\ &\geq \frac{\delta}{2(C+\delta)} \\ &> 0 \end{aligned}$$

□

Wolfowitz shows that one can even prove $P_e \to 1$, see also Ahlswede's proof, but this requires a different type of typicality which we will not pursue here.