Information Theory for Wireless Communications. Lecture 4

Instructor: Dr. Saif Khan Mohammed
Scribe: Antonios Pitarokoilis

I. Strong Converse of the Channel Coding Theorem

Theorem 1: *Strong Converse of the Channel Coding Theorem*

Let $\epsilon > 0$ and $0 \leq \lambda < 1$ be arbitrary. For n sufficiently large there does *not* exist a $(n, 2^{n(C+\epsilon)}, \lambda)$ code.

We will prove the following stronger result which implies Theorem 1.

Theorem 2: Let $0 \leq \lambda < 1$ be arbitrary. There exists positive constant K' such that for *any* n, there does not exist a $(n, 2^{n(C+\frac{K'}{\sqrt{n}})}, \lambda)$ code. K' depends on $\lambda,\ M$ and M'.

Lemma 1: Let $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ be any (n, N, λ) code such that u_i are K_0-input typical and $A_i \subset \mathcal{B}_2^{\delta,K_0}(u_i),\ \forall i$. Then $N < 2^{n\left(H(Y)-H(Y|X)+\frac{K_8(\lambda,\delta,K_0)}{\sqrt{n}}\right)}$ for some constant $K_8(\lambda, \delta, K_0) > 0$.

Proof: Since $A_i \subset \mathcal{B}_2^{\delta,K_0}(u_i)$, $1 > P\left(Y^n \in A_i | X^n = u_i\right) > 1-\lambda$. Additionally, from eq. 7 in Lemma 1 in [1], we have that for any $y^n \in \mathcal{B}_2^{\delta,K_0}(u_i)$,

$$P\left(Y^n = y^n | X^n = u_i\right) < 2^{-n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)}. \tag{1}$$

Therefore,

$$1-\lambda < P\left(Y^n \in A_i | X^n = u_i\right) = \sum_{y^n \in A_i} P(Y^n = y^n | X^n = u_i) < |A_i|\, 2^{-n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)} \Rightarrow$$
$$|A_i| > (1-\lambda)2^{n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)}$$

Due to the fact that the A_i's are disjoint we have

$$\left|\cup_{i=1}^{N} A_i\right| > N(1-\lambda)2^{n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)}. \tag{2}$$

We repeat the definition of the set $\mathcal{B}_1^{\delta,K_0} \triangleq \{y^n \in \mathcal{Y}^n | y^n$ is $\delta-$ generated by some K_0- input typical$\}$.

Therefore, we have

$$\cup_{i=1}^{N} A_i \subseteq \mathcal{B}_1^{\delta,K_0} \Rightarrow \left|\cup_{i=1}^{N} A_i\right| \leq \left|\mathcal{B}_1^{\delta,K_0}\right| \tag{3}$$

From Lemma 6 in [2] we get an upper bound on the size of the set $\mathcal{B}_1^{\delta,K_0}$, i.e.

$$\left|\mathcal{B}_1^{\delta,K_0}\right| < 2^{n\left(H(Y)+\frac{K_4}{\sqrt{n}}\right)}. \tag{4}$$

Combining (2), (3) and (4) we have

$$N(1-\lambda)2^{n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)} < \left|\cup_{i=1}^{N} A_i\right| < 2^{n\left(H(Y)+\frac{K_4}{\sqrt{n}}\right)}$$

$$N < \frac{1}{1-\lambda} 2^{n(I(X;Y)+\frac{K_4+K_5}{\sqrt{n}})} = 2^{n(I(X;Y)+\frac{K_4+K_5}{\sqrt{n}}+\frac{\log\frac{1}{1-\lambda}}{n})} < 2^{n\left(I(X;Y)+\frac{K_8}{\sqrt{n}}\right)},$$

where $K_8(\lambda,\delta,K_0) \triangleq K_4 + K_5 + \log\frac{1}{1-\lambda}$. In the last step we have used the fact that $\frac{1}{n}\log\frac{1}{1-\lambda} < \frac{1}{\sqrt{n}}\log\frac{1}{1-\lambda},\ n \geq 1$. ■

Lemma 2: Let $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ be a (n, N, λ) code such that u_i are K_0-input typical.[1] Then,

$$N < 2^{n\left(I(X;Y)+\frac{K_9(\lambda,K_0)}{\sqrt{n}}\right)},$$

for some constant $K_9(\lambda, K_0) > 0$.

Proof: Choose δ in Lemma 2 in [2] be such that $\delta^2 > \frac{2MM'}{1-\lambda}$, i.e.

$$P\left(Y^n \in \mathcal{B}_2^{\delta,K_0}(x^n) | X^n = x^n\right) > 1 - \frac{1-\lambda}{2}.$$

Define a new code $\{(u_1, A_1'), \ldots, (u_N, A_N')\}$, where $A_i' \triangleq A_i \cap \mathcal{B}_2^{\delta,K_0}(u_i)$.

$$P\left(Y^n \in A_i' | X^n = u_i\right) = P\left(Y^n \in A_i | X^n = u_i\right) + P\left(Y^n \in \mathcal{B}_2^{\delta,K_0} | X^n = u_i\right)$$
$$- P\left(Y^n \in A_i \cup \mathcal{B}_2^{\delta,K_0} | X^n = u_i\right) > (1-\lambda) + \left(1 - \frac{1-\lambda}{2}\right) - 1 = \frac{1-\lambda}{2}.$$

The new code is $(n, N, 1-\frac{1-\lambda}{2})$ and therefore from Lemma 1, it follows that $N < 2^{n\left(I(X;Y)+\frac{K_9(\lambda,K_0)}{\sqrt{n}}\right)}$, where $K_9(\lambda,K_0) \triangleq K_8(\frac{1+\lambda}{2},\delta,K_0)$ for $\delta^2 > \frac{2MM'}{1-\lambda}$. ■

Before stating the proof of Theorem 2 we define the empirical probability density function of any input

[1]We note that A_i need *not* be a subset of $\mathcal{B}_2^{\delta,K_0}(u_i)$. For the achievability proof of the channel coding theorem for discrete memoryless channels, we used a constructive proof to get a code that achieves a rate arbitrarily close to the channel capacity. In that proof A_i was selected to be a subset of $\mathcal{B}_2^{\delta,K_0}(u_i)$. However, there might be codes that achieve higher rates and do not meet this constraint in terms of the A_i's. Here in the converse we prove that there is no code that can achieve rates higher that the capacity. Therefore, we have to include all the possible codes and not only those that result from the constructive procedure followed for the achievability proof.

sequence $x^n \in \mathcal{X}^n$, which will be useful later for the proof of Theorem 2.

Definition 1: Consider any $x^n \in \mathcal{X}^n$. We define the empirical probability density function (pdf) of a given sequence x^n as $p_{x^n}(i) \stackrel{\Delta}{=} \frac{f_i(x^n)}{n}$, where $f_i(x^n)$ is the number of locations in the sequence x^n where letter $a_i,\ i = 1, \ldots, M$, appears.

Example 1: Consider the binary $(M = 2)$ alphabet $\mathcal{X} = \{a_1 = 0, a_2 = 1\}$. The empirical pdf of the given sequence $x^n = 010001011110000$ is $\{p_{x^n}(a_1), p_{x^n}(a_2)\} = \{9/15, 6/15\}$.

Remark 1: Note that x^n is K_0-input typical with respect to its own empirical pdf for any $K_0 > 2$.

Proof: (Theorem 2) Define a set $\mathfrak{P}^{(n)} \stackrel{\Delta}{=} \{\left(\frac{\alpha_1}{n}, \ldots, \frac{\alpha_M}{n}\right) | \alpha_i \in \{0, 1, 2, \ldots, n\}$ and $\sum_{i=1}^{M} \alpha_i = n\}$. Since for $\mathfrak{S}^{(n)} \stackrel{\Delta}{=} \{\left(\frac{\alpha_1}{n}, \ldots, \frac{\alpha_M}{n}\right) | \alpha_i \in \{0, 1, 2, \ldots, n\}\}$ it holds $\left|\mathfrak{S}^{(n)}\right| = (n+1)^M$ and $\mathfrak{P}^{(n)} \subset \mathfrak{S}^{(n)}$ we have $\left|\mathfrak{P}^{(n)}\right| < (n+1)^M$.

Let $\mathcal{C} = \{(u_1, A_1), \ldots, (u_N, A_N)\}$, which is a (n, N, λ) code as defined in [2]. For this code, u_i need not be K_0-input typical with respect to some discrete memoryless source, nor need the A_i be the set of sequences that are $\delta-$generated by a K_0-input typical sequence u_i. Based on the previous considerations, we split the code $\mathcal{C}$ into a finite number of subcodes. This is done by splitting the codewords of $\mathcal{C}$ into K different subgroups, $\mathcal{G}_l = \{u_{l_1}, \ldots, u_{|\mathcal{G}_l|}\},\ l = 1, \ldots, K$, in such a way that the codewords in the $l-$th group are K_0-input typical to a pdf, $\Pi_l \in \mathfrak{P}^{(n)}$. In the following we present the splitting procedure in more detail. We start by selecting the first codeword u_1. This codeword has an empirical pdf, p_{u_1}, and we define $\Pi_1 \equiv p_{u_1}$. The next codeword, u_2, can either be placed into the first group if it is K_0-input typical with respect to Π_1 or else we create a new group for the codewords that are K_0-input typical to p_{u_2} and we define as $\Pi_2 \equiv p_{u_2}$. At the $i-$th step, $g(i) \leq i$ groups have been created, where each codeword in the $l-$th group, $l \leq g(i)$ is K_0-input typical to the pdf $\Pi_l \in \mathfrak{P}^{(n)}$. If the codeword u_{i+1} is K_0-input typical with respect to some pdf $\Pi_l,\ (l = \{1, \ldots, g(i)\})$, then u_{i+1} is placed into the corresponding group, else a new group is created which contains the codewords that are K_0-input typical with respect to $p_{u_{i+1}} \equiv \Pi_{g(i)+1}$. This procedure stops when all the N codewords, which form a finite set, have been placed into some group $\mathcal{G}_l$. Therefore, the number of the created subcodes, $K \leq N$, is also finite. Since $\{\Pi_1, \ldots, \Pi_K\} \subset \mathfrak{P}^{(n)}$ it follows that $K < (n+1)^M$.

The groups formed by the previous procedure create K subcodes of the original code. The $l-$th subcode, given by $\mathcal{C}_l = \{(u_{l_1}, A_{l_1}), \ldots, (u_{l_{|\mathcal{G}_l|}}, A_{l_{|\mathcal{G}_l|}})\}$ is a $(n, |\mathcal{G}_l|, \lambda)$ code with each codeword being K_0-input

typical to Π_l. For each code $\mathcal{C}_l$, by Lemma 2 we have a bound on the number of codewords in $\mathcal{C}_l$, i.e.

$$|\mathcal{G}_l| < 2^{n\left(I_{\Pi_l}(X;Y)+\frac{1}{\sqrt{n}}K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)\right)} \leq 2^{n\left(C+\frac{1}{\sqrt{n}}K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)\right)}, \tag{5}$$

where $\epsilon > 0$ is some arbitrary constant. We have also used the fact that $I_{\Pi_l}(X;Y) \leq C$. Using (5), we can obtain a bound on the number of codewords in the initial code, $\mathcal{C}$.

$$\begin{aligned} N = \sum_{l=1}^{K} |\mathcal{G}_l| &< K\ 2^{n\left(C+\frac{1}{\sqrt{n}}K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)\right)} < (n+1)^M\ 2^{n\left(C+\frac{1}{\sqrt{n}}K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)\right)} \\ &= 2^{n\left(C+\frac{1}{\sqrt{n}}K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)+\frac{M\log_2(n+1)}{n}\right)} < 2^{n\left(C+\frac{1}{\sqrt{n}}\left(K_8\left(\frac{1+\lambda}{2},\sqrt{\frac{2MM'}{1-\lambda}}+\epsilon,K_0\right)+4M\right)\right)}, \end{aligned}$$

where we have used the fact that $\log_2(n+1) < 4\sqrt{n}$, for $n \geq 1$. This then concludes the proof. ■

II. Weak Converse to the Channel Coding Theorem

Before stating the proof of the weak converse of the channel coding theorem, we introduce the necessary notation. We consider sequences x^n of length n, where each symbol in the sequence is chosen from an alphabet $\mathcal{X}$, i.e. $x^n \in \mathcal{X}^n$. On these sequences we define a probability mass function $Q'(\cdot)$, such that $\sum_{x^n\in\mathcal{X}^n} Q'(x^n) = 1$. Let the channel probability function be denoted by $h(Y^n = y^n|X^n = x^n)$, where $h(Y^n = y^n|X^n = x^n) = \prod_{i=1}^{n} P\left(Y_i = y_i|X_i = x_i\right)$, which follows from the fact that we consider the Discrete Memoryless Channel (DMC). Let $U \triangleq X^n$ and $V \triangleq Y^n$ be the input and output sequences of the DMC, respectively. We denote the joint distribution of U and V as $Q(U = x^n, V = y^n) = Q(x^n)h(y^n|x^n)$. The marginal distribution of the output V is denoted by $Q''(V = y^n) = \sum_{x^n\in\mathcal{X}^n} Q'(x^n)h(y^n|x^n)$. Also, let $Q(U,V)$ and $Q'(U)Q''(V)$, $I(U;V) \triangleq \log_2\left(\frac{Q(U,V)}{Q'(U)Q''(V)}\right)$. We define the random variable $J_{Q'} \triangleq I(U;V)$ and the number $R_{Q'} \triangleq \mathbb{E}_{U,V}\left[J_{Q'}\right]$. For sake of continuity we repeat the definition of a code. A (n, N, λ) code is defined as a set $\{(u_1, A_1), \ldots, (u_N, A_N)\}$, such that

1) $u_i \in \mathcal{X}^n,\ A_i \subset \mathcal{Y}^n$
2) $A_i \cap A_j = \emptyset,\ i \neq j$ and $\cup_{i=1}^{N} A_i = \mathcal{Y}^n$ without loss of generality
3) $\sum_{y^n\in A_i} h(y^n|u_i) > 1-\lambda,\ i = 1, \ldots, N$ and
4) $N \geq 2$

Theorem 3: Let $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ be an (n, N, λ) code. We define the following input

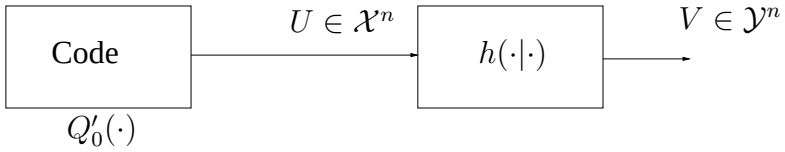

Fig. 1. System model for Theorem 3.

distribution on U

$$Q_0'(U = x^n) = \begin{cases} 1/N & \text{if } x^n \in (u_1, \ldots, u_N) \\ 0 & \text{otherwise.} \end{cases}$$

For the system in Fig. 1 it holds

$$H(U|V) \leq \bar{\lambda} \log_2(N-1) - \bar{\lambda} \log_2 \bar{\lambda} - (1-\bar{\lambda}) \log_2(1-\bar{\lambda}) \tag{6}$$

where $\bar{\lambda} \stackrel{\Delta}{=} \frac{1}{N} \sum_{i=1}^{N} \lambda(u_i)$ and $\lambda(u_i) \stackrel{\Delta}{=} \sum_{y^n \notin A_i} h(y^n|u_i)$.

Proof: Consider the random variable

$$V' \stackrel{\Delta}{=} \{i, \text{ if } V \in A_i\} = f(V),$$

which is a deterministic function, f, of V. Therefore, it follows that

$$H(U|V) \leq H(U|V') = \sum_{j=1}^{N} P(V'=j) H(U|V'=j), \tag{7}$$

where

$$\begin{aligned} H(U|V'=j) &= -\sum_{i=1}^{N} P(U=u_i|V'=j) \log_2 P(U=u_i|V'=j) \\ &= -P(U=u_j|V'=j) \log_2 P(U=u_j|V'=j) \\ &\quad -\sum_{\substack{i=1 \\ i \neq j}}^{N} P(U=u_i|V'=j) \log_2 P(U=u_i|V'=j) \end{aligned} \tag{8}$$

Consider the distribution

$$q_i \stackrel{\Delta}{=} \frac{P(U=u_i|V'=j)}{\sum_{\substack{k=1 \\ k \neq j}}^{N} P(U=u_i|V'=j)}, \ i = 1, \ldots, N, \ i \neq j.$$

Due to the fact that the uniform distribution maximizes the entropy over all the probability distributions with finite support, it follows that

$$-\sum_{\substack{i=1\\ i\neq j}}^{N} q_i \log_2 q_i \leq \log_2(N-1). \tag{9}$$

Further, we also have $P(U \neq u_j|V'=j) = \sum_{\substack{i=1\\ i\neq j}}^{N} P(U=u_i|V'=j)$. It holds

$$\begin{aligned}
&-\sum_{\substack{i=1\\ i\neq j}}^{N} P(U=u_i|V'=j)\log_2 P(U=u_i|V'=j)\\
=&-\sum_{\substack{i=1\\ i\neq j}}^{N} \frac{P(U=u_i|V'=j)P(U\neq u_j|V'=j)}{P(U\neq u_j|V'=j)} \log_2 \frac{P(U=u_i|V'=j)P(U\neq u_j|V'=j)}{P(U\neq u_j|V'=j)}\\
=&-\sum_{\substack{i=1\\ i\neq j}}^{N} q_i P(U\neq u_j|V'=j)\log_2 q_i P(U\neq u_j|V'=j)\\
=&-P(U\neq u_j|V'=j)\sum_{\substack{i=1\\ i\neq j}}^{N} q_i\log_2 q_i - P(U\neq u_j|V'=j)\log_2 P(U\neq u_j|V'=j)\sum_{\substack{i=1\\ i\neq j}}^{N} q_i\\
\leq& P(U\neq u_j|V'=j)\log_2(N-1) - P(U\neq u_j|V'=j)\log_2 P(U\neq u_j|V'=j).
\end{aligned} \tag{10}$$

The last equality follows from (9) and the fact that $\sum_{\substack{i=1\\ i\neq j}}^{N} q_i = 1$. Therefore,

$$\begin{aligned}
H(U|V'=j) \leq& -P(U=u_j|V'=j)\log_2 P(U=u_j|V'=j) + P(U\neq u_j|V'=j)\log_2(N-1)\\
&-P(U\neq u_j|V'=j)\log_2 P(U\neq u_j|V'=j).
\end{aligned} \tag{11}$$

Define

$$Z(U,V') \triangleq \begin{cases} 1 & \text{if } U=u_i, V'=i \text{ for some } i\in\{1,\ldots,N\}\\ 0 & \text{otherwise.}\end{cases}$$

Therefore, $Z(U,V')=1 \Leftrightarrow (U,V')\in\{(u_1,1),\ldots,(u_N,N)\}$. Further,

$$\begin{aligned}
P(Z=0) = 1-P(Z=1) &= 1-\sum_{i=1}^{N} P(U=u_i,V'=i) = 1-\sum_{i=1}^{N} P(V'=i|U=u_i)P(U=u_i)\\
&= 1-\frac{1}{N}\sum_{i=1}^{N}(1-\lambda(u_i)) = \bar{\lambda}
\end{aligned}$$

Using (11) we have

$$\begin{aligned}
H(U|V') \leq & -\sum_{j=1}^{N} \left(P(U=u_j|V'=j)\log_2 P(U=u_j|V'=j)\right) P(V'=j) \\
& -\sum_{j=1}^{N} \left(P(U\neq u_j|V'=j)\log_2 P(U\neq u_j|V'=j)\right) P(V'=j) \\
& +\log_2(N-1)\sum_{j=1}^{N} P(U\neq u_j|V'=j)P(V'=j) \\
\leq & -\sum_{j=1}^{N} \underbrace{P(U=u_j,V'=j)}_{P(Z=1,V'=j)} \log_2 \underbrace{P(U=u_j|V'=j)}_{P(Z=1|V'=j)} \\
& -\sum_{j=1}^{N} \underbrace{P(U\neq u_j,V'=j)}_{P(Z=0,V'=j)} \log_2 \underbrace{P(U\neq u_j|V'=j)}_{P(Z=0|V'=j)} \\
& +\log_2(N-1)\sum_{j=1}^{N} P(Z=0,V'=j) = H(Z|V') + \bar{\lambda}\log_2(N-1),
\end{aligned}$$

where the last equality follows by the fact that $\sum_{j=1}^{N} P(Z=0,V'=j) = P(Z=0) = \bar{\lambda}$. Finally,

$$\begin{aligned}
H(U|V) \leq & H(U|V') \leq H(Z|V') + \bar{\lambda}\log_2(N-1) \leq H(Z) + \bar{\lambda}\log_2(N-1) \\
& \leq -\bar{\lambda}\log_2\bar{\lambda} - (1-\bar{\lambda})\log_2(1-\bar{\lambda}) + \bar{\lambda}\log_2(N-1) \leq 1 + \bar{\lambda}\log_2(N-1).
\end{aligned}$$

■

REFERENCES

[1] H. Q. Ngo, "Information Theory for Wireless Communication, Lecture 3:Conditional Typicality, Channel Coding Theorem for DMC".

[2] T. V. K. Chaitanya, "Information Theory for Wireless Communication, Lecture 2:Conditionally and Jointly Typical Sequences".