Information Theory for Wireless Communication

Lecture 3: Conditional Typicality, Channel Coding Theorem for DMC

Lecture by Dr. Saif K. Mohammed

Scribe by Hien Quoc Ngo

In this lecture, we introduce conditional typicality, some properties of the entropy function, and prove the channel coding theorem (achievability).

I. Conditional Typicality

Definition 1. Let x^n be a K_0-input typical sequence. We define the conditional typicality set $B_2^{(\delta,k_0)}(x^n)$ as

$$B_2^{(\delta,k_0)}(x^n) = \{y^n \in \mathcal{Y}^n | y^n \text{is } \delta\text{-generated by } x^n\}. \tag{1}$$

Lemma 1. *Suppose* $B_2^{(\delta,k_0)}(x^n)$ *is defined as* (1)*, we have*

$$2^{n\left(H(Y|X)-\frac{K_6}{\sqrt{n}}\right)} < \left|B_2^{(\delta,k_0)}(x^n)\right| < 2^{n\left(H(Y|X)+\frac{K_6}{\sqrt{n}}\right)} \tag{2}$$

where $K_6 \triangleq 4MM'\sqrt{MK_0}(1+\delta)\frac{\log_2 e}{e} - log_2\left(1-\frac{MM'}{\delta^2}\right)$.

Proof: We have $\Pr(Y^n = y^n | X^n = x^n) = \prod_{i=1}^{M}\prod_{j=1}^{M'} p(j|i)^{f_{i,j}(x^n,y^n)}$. Therefore

$$-\frac{\log \Pr(Y^n = y^n | X^n = x^n)}{n} = \sum_{i=1}^{M}\sum_{j=1}^{M'} \frac{f_{i,j}(x^n,y^n)}{n} \log \frac{1}{p(j|i)}. \tag{3}$$

For any $y^n \in B_2^{(\delta,k_0)}(x^n)$, it is also true that $(x^n, y^n) \in B^{(\delta,k_0)}$ (see Definition 2.3 in Lecture 2 [1]). From the proof of Lemma 2.4[1], we have

$$\frac{f_{i,j}(x^n,y^n)}{n} < p(i,j) + \frac{\sqrt{MK_0}}{\sqrt{n}}(1+\delta)p(i)^{1/4}p(j|i)^{1/4} \tag{4}$$

[1]In Lecture 2, we had only stated this Lemma, with the proof being one of the questions in Problem set I.

Substituting (4) into (3), we obtain

$$-\frac{\log \Pr\left(Y^n = y^n | X^n = x^n\right)}{n} < \underbrace{\sum_{i=1}^{M}\sum_{j=1}^{M'} p\left(i\right) p\left(j|i\right) \log \frac{1}{p\left(j|i\right)}}_{H(Y|X)} + \frac{K_5}{\sqrt{n}} \tag{5}$$

where $K_5 \triangleq 4MM'\sqrt{MK_0}\left(1+\delta\right)\frac{\log_2 e}{e}$. Therefore

$$\Pr\left(Y^n = y^n | X^n = x^n\right) > 2^{-n\left(H(Y|X)+\frac{K_5}{\sqrt{n}}\right)} \tag{6}$$

Similarly, it can also be shown that

$$\Pr\left(Y^n = y^n | X^n = x^n\right) < 2^{-n\left(H(Y|X)-\frac{K_5}{\sqrt{n}}\right)} \tag{7}$$

Furthermore, we can apply Lemma 2.2 to obtain

$$1 - \frac{MM'}{\delta^2} < \Pr\left(Y^n \in B_2^{(\delta,k_0)}\left(x^n\right) | x^n \text{was transmitted}\right) < 1 \tag{8}$$

From (6), (7), (8), and using the technique described earlier for bounding the cardinality of a given set, we arrive at the desired result (2).[2] ■

Essentially $B_2^{(\delta,k_0)}\left(x^n\right)$ is a high-probability set for the channel output sequence conditioned on the input being the K_0-input typical sequence x^n. In Lemma 1 we have shown that the size of this high probability set increases exponetially with n as $2^{n(H(Y|X)+O(1/\sqrt{n}))}$, and therefore $H(Y|X)$ is referred to as the entropy of Y conditioned on X.

II. Basic Properties of Entropy Function

Entropy represents the uncertainty of a random variable. Let X and Y be discrete random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $p(x)$, $p(x,y)$, and $p(y|x)$ be the probability mass function of X, joint distribution, and conditional distribution, respectively. Then,

- The entropy $H(X)$ of X is defined by

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log\left(p(x)\right). \tag{9}$$

[2]The set size bounding technique is discussed on page 4, Lecture 2 notes [1].

- The joint entropy $H(X,Y)$ of a pair of discrete random variables (X,Y) is defined by

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left(p(x,y)\right). \tag{10}$$

- The conditional entropy $H(Y|X)$ is defined by

$$H(Y|X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left(p(y|x)\right). \tag{11}$$

Some main properties of entropy:

- $H(X) > 0$.
- $H(p)$ is concave in p.
- Chain rule:

$$H(X_1, X_2, ..., X_n) = H(X_1) + H(X_2|X_1) + ... + H(X_n|X_1, ..., X_{n-1}). \tag{12}$$

- For any two random variables, X and Y, we have

$$H(X) \geq H(X|Y) \tag{13}$$

  with equality if and only if X and Y are independent. The above property implies that conditioning reduces entropy.
- $H(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$, with equality if and only if the X_i are independent.
- Let $\Phi : \mathcal{X} \to \mathcal{Y}$ be a deterministic function. Define $Y = \Phi(X)$. Then

$$H(Y) \leq H(X).$$

III. Channel Coding Theorem for DMC

Theorem 2. *Let $0 < \lambda \leq 1$ be any arbitrary number, there exists a positive scalar constant $k(\lambda)$ such that for any given n, there exists a code with parameters (n, N, λ) with*[3]

$$N > 2^{n\left(C - \frac{k(\lambda)}{\sqrt{n}}\right)} \tag{14}$$

where $C = \max_{\{p_i\}}(H(Y) - H(Y|X))$.

Proof: The proof is by construction. We construct a code which satisfies (14).

[3]The definition of a code (n, N, λ) can be found in Section. II of previous lecture [1].

Code construction:

Choose $K_0 > 2$. Choose $\delta^2 > MM'$ such that $\frac{MM'}{\delta^2} < \frac{\lambda}{2}$, i.e., $\delta^2 > \frac{2MM'}{\lambda}$. We can now construct a (n, N, λ) code which satisfies the following:

- u_i is a K_0-input typical sequence, for $i = 1, 2, ..., N$.
- $A_i = B_2^{(\delta,k_0)}(u_i) \ - \ \{A_1 \cup A_2 \cup ... \cup A_{i-1}\}$, $i = 1, 2, ..., N$.
- It is impossible to extend this (n, N, λ) code to a $(n, N+1, \lambda)$ code by adding another (codeword,decoding set) pair (u_{N+1}, A_{N+1}) which satisfies the above two conditions.

We now prove that the code constructed above satisfies (14).

Step 1: We show that for any $x^n \in \mathcal{T}_{K_0}$, $\Pr(Y^n \in A_1 \cup ... \cup A_N | x^n \text{was transmitted}) > \frac{\lambda}{2}$:

Since $\delta^2 > \frac{2MM'}{\lambda}$, from Lemma 2.2, we have

$$\Pr(Y^n \text{is not } \delta\text{-generated by } x^n | X^n = x^n) < \frac{\lambda}{2}. \tag{15}$$

▷ If $x^n = u_i$. Then

$$\begin{aligned}\Pr(Y^n \in A_1 \cup ... \cup A_N | x^n \text{was transmitted}) &> \Pr\left(Y^n \in B_2^{(\delta,k_0)}(x^n) | X^n = x^n\right) \\ &> 1 - \frac{\lambda}{2} \geq \frac{\lambda}{2}.\end{aligned} \tag{16}$$

where the first inequality is obtained from the fact that $B_2^{(\delta,k_0)}(u_i) \subseteq A_1 \cup ... \cup A_N$, and the second inequality is obtained from (15).

▷ If $x^n \in \mathcal{T}_{K_0}$, and $x^n \notin \{u_1, ..., u_N\}$. Then, we have

$$\begin{aligned}&\Pr(Y^n \in A_1 \cup ... \cup A_N | x^n \text{was transmitted}) \\ &> \Pr\left(Y^n \in B_2^{(\delta,k_0)}(x^n) \cap (A_1 \cup ... \cup A_N) | X^n = x^n\right) \\ &= \underbrace{\Pr\left(Y^n \in B_2^{(\delta,k_0)}(x^n) | X^n = x^n\right)}_{>1-\lambda/2 \text{ (a)}} - \underbrace{\Pr\left(Y^n \in B_2^{(\delta,k_0)}(x^n) \ - \ (A_1 \cup ... \cup A_N) | X^n = x^n\right)}_{<1-\lambda \text{ (b)}} \\ &> \frac{\lambda}{2}.\end{aligned} \tag{17}$$

where (a) follows from (15) and (b) follows from the fact that the code cannot be extended to a $(n, N+1, \lambda)$ code (i.e., if any $x^n \in \mathcal{T}_{K_0}$ is used as the $N+1$-th codeword, then the error probability given that x^n is transmitted is the probability that $Y^n \notin B_2^{(\delta,k_0)}(x^n) \ - \ (A_1 \cup ... \cup A_N)$. This error

probability must be greater than λ since otherwise we can add x^n as the $N+1$-th codeword and extend our code.).

Step 2: Next we show that $\Pr\left(Y^n \in A_1 \cup ... \cup A_N\right) > \left(1 - 1/K_0\right)\frac{\lambda}{2}$. We have

$$\begin{aligned}
\Pr\left(Y^n \in A_1 \cup ... \cup A_N\right) &= \sum_{y^n \in A_1 \cup ... \cup A_N} \Pr\left(Y^n = y^n\right) \\
&= \sum_{y^n \in A_1 \cup ... \cup A_N} \sum_{x^n \in \mathcal{X}^n} \Pr\left(Y^n = y^n, X^n = x^n\right) \\
&= \sum_{x^n \in \mathcal{X}^n} \Pr\left(X^n = x^n\right) \sum_{y^n \in A_1 \cup ... \cup A_N} \Pr\left(Y^n = y^n | X^n = x^n\right) \\
&> \sum_{x^n \in \mathcal{T}_{K_0}} \Pr\left(X^n = x^n\right) \sum_{y^n \in A_1 \cup ... \cup A_N} \Pr\left(Y^n = y^n | X^n = x^n\right) \\
&\underset{>}{(a)} \sum_{x^n \in \mathcal{T}_{K_0}} \Pr\left(X^n = x^n\right) \frac{\lambda}{2} \\
&\underset{>}{(b)} \left(1 - 1/K_0\right) \frac{\lambda}{2}
\end{aligned} \tag{18}$$

where in (a), we have used the result of Step 1, and in (b), we have used Lemma 2.1.

Step 3: From the result in Step 2, we have

$$\begin{aligned}
\left(1 - 1/K_0\right) \frac{\lambda}{2} &< \Pr\left(Y^n \in A_1 \cup ... \cup A_N\right) \\
&= \sum_{y^n \in A_1 \cup ... \cup A_N} \Pr\left(Y^n = y^n\right) < |A_1 \cup ... \cup A_N| 2^{-n\left(H(Y) - \frac{K_3}{\sqrt{n}}\right)}
\end{aligned} \tag{19}$$

where the last inequality is obtained using equation (22) in the proof of Lemma 2.6 (since $A_1 \cup ... \cup A_N \subset B_1^{(\delta, k_0)}$). Therefore,

$$|A_1 \cup ... \cup A_N| > \left(1 - 1/K_0\right) \frac{\lambda}{2} 2^{n\left(H(Y) - \frac{K_3}{\sqrt{n}}\right)} \tag{20}$$

Furthermore, since $A_i \in B_2^{(\delta, k_0)}\left(u_i\right)$, and $|B_2^{(\delta, k_0)}\left(u_i\right)| \leq 2^{n\left(H(Y|X) + \frac{K_6}{\sqrt{n}}\right)}$, we have

$$|A_1 \cup ... \cup A_N| = \sum_{i=1}^{N} |A_i| < N 2^{n\left(H(Y|X) + \frac{K_6}{\sqrt{n}}\right)} \tag{21}$$

From (20) and (21), we obtain

$$N > 2^{n\left(I(X;Y) - \frac{K_7}{\sqrt{n}}\right)} \tag{22}$$

where $K_7 \triangleq K_3 + K_6 - \log\left(\frac{\lambda}{2}\left(1 - 1/K_0\right)\right)$ and $I(X;Y) \triangleq H(Y) - H(Y|X)$. Since (22) is true for all

input distributions, it is also true for the input distribution which maximizes $I(X;Y)$. This then completes the proof. ■

REFERENCES

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

[2] J. Wolfowitz, *Coding Theorems of Information Theory*, Ergebnisse Der Mathematik Und Ihrer Grenzgebiete, 2nd Ed. 1964.