Information Theory for Wireless Communications: Lecture 5

Instructor: Dr. Saif K. Mohammed
Scribe: Johannes Lindblom

I. Proof of the Weak Converse to the Channel Coding Theorem of the DMC

Let $u \in \mathcal{X}^n$ and $v \in \mathcal{Y}^n$ be the random variables representing the input and output of the DMC with probability density functions $Q'(\cdot)$ and $Q''(\cdot)$, respectively. Let x^n be a input to a DMC with output y^n. Then, we define

$$h(y^n|x^n) \triangleq \Pr\{v = y^n | u = x^n\} = \prod_{i=1}^{n} p(y_i|x_i),$$

where $p(\cdot|\cdot)$ is the c.p.f. of the DMC. Also

$$Q(u = x^n, v = y^n) \triangleq \Pr\{u = x^n, v = y^n\} = Q'(x^n)h(y^n|x^n)$$

$$Q''(v = y^n) = \sum_{x^n \in \mathcal{X}^n} Q(u = x^n, v = y^n).$$

We define

$$I(x^n, y^n) \triangleq \log_2 \frac{Q(x^n, y^n)}{Q'(x^n)Q''(y^n)},$$

$$J(Q') \triangleq I(u, v), \text{ and } R(Q') \triangleq \mathrm{E}_{u,v}[I(u, v)].$$

Theorem 1. *Let $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ be a (n, N, λ) code. We define the input distribution*

$$Q'_0(u = x^n) = \begin{cases} 1/N, & x^n \in \{u_1, \ldots, u_N\} \\ 0, & \textit{otherwise.} \end{cases}$$

Let u be the input to a channel with output v. Then,

$$H(u|v) \leq \bar{\lambda} \log(N - 1) - \bar{\lambda} \log \bar{\lambda} - (1 - \bar{\lambda}) \log(1 - \bar{\lambda}),$$

where

$$\bar{\lambda} \triangleq \frac{1}{N} \sum_{i=1}^{N} \lambda(u_i)$$

with

$$\lambda(u_i) = \sum_{y^n \notin A_i} h(y^n|u_i)$$

Proof: See [1]. ■

We define

$$T \triangleq \max_{Q'(\cdot)} R(Q') \tag{1}$$

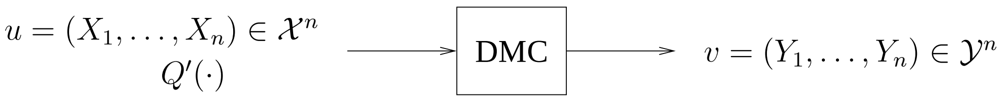

Fig. 1. System model for Theorem 3.

Theorem 2. *Any* (n, N, λ) *channel code satisfies*

$$\log N \leq \frac{T+1}{1-\lambda}$$

Proof: From (1), we see that $T \geq R(Q'_0)$. We have

$$\begin{aligned} R(Q'_0) &= H(u) - H(u|v) = \log N - H(u|v) \\ &\geq \log N - \bar{\lambda}\log(N-1) + \bar{\lambda}\log\bar{\lambda} + (1-\bar{\lambda})\log(1-\bar{\lambda}), \end{aligned}$$

where the inequality follows from Theorem 1. Then it follows that

$$\begin{aligned} T + 1 &\geq R(Q'_0) + 1 \\ &\geq \log N - \bar{\lambda}\log(N-1) + \underbrace{\bar{\lambda}\log\bar{\lambda} + (1-\bar{\lambda})\log(1-\bar{\lambda}) + 1}_{\geq 0} \\ &\geq \log N - \bar{\lambda}\log(N-1) \geq (1-\bar{\lambda})\log N, \end{aligned}$$

where the last inequality follows from $-\log(N-1) > -\log(N)$. Since $\bar{\lambda} \leq \lambda$, we have $1 - \bar{\lambda} \geq 1 - \lambda$ and

$$T + 1 \geq (1-\lambda)\log N \Leftrightarrow \log N \leq \frac{T+1}{1-\lambda}. \tag{2}$$

■

Theorem 3. *For the DMC we defined in Fig. 1, we have*

$$T = \max_{Q'} R(Q') = nC,$$

where

$$C \triangleq \max_{\{p(i)\}} \sum_i \sum_j p(i,j) \log \frac{p(i,j)}{p(i)p(j)}$$

Proof: We have

$$\begin{aligned} R(Q') &= H(v) - H(v|u) = H(Y_1, \ldots, Y_n) - H(Y_1, \ldots, Y_n | X_1, \ldots, X_n) \\ &\overset{(a)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \overset{(b)}{\leq} nC \end{aligned}$$

Inequality (a) is tight when the Y_i's are independent. Inequality (b) is tight for the optimal input distribution.

■

From Theorem 3 and (2), we get

$$\log N \leq \frac{nC+1}{1-\lambda}.$$

Given *any arbitrary* constant $\epsilon > 0$, for a $(n, 2^{n(C+\epsilon)}, \lambda)$ code to exist, we must have

$$\log 2^{n(C+\epsilon)} \leq \frac{nC+1}{1-\lambda}. \Leftrightarrow \lambda \geq \frac{\epsilon}{C+\epsilon} - \frac{1}{n(C+\epsilon)}$$

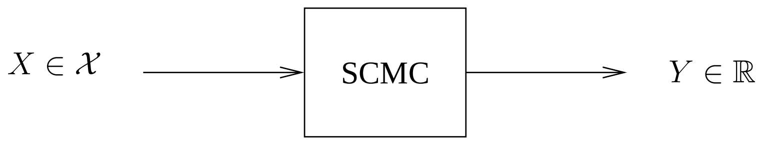

Fig. 2. System model used in Sec. II.

Hence, for $\lambda < \epsilon/(C+\epsilon)$ and n is sufficiently large, no $(n, 2^{n(C+\epsilon)}, \lambda)$ code can exist. This is the weak converse.

II. Capacity of the Discrete-Time Semi-Continuous Memoryless Channel (SCMC)

Consider the system depicted in Fig. 2 where the input is discrete and the output is continuous valued. Let $f_{Y|i}(y)$ be the p.d.f. of the output Y given that the input is $a_i \in \mathcal{X}$. The cumulative density function of the output is

$$F_{Y|i}(y) \triangleq \Pr\{Y \leq y | a_i\} = \int_{-\infty}^{y} f_{Y|i}(t) dt$$

$$f_{Y|i}(y) = \frac{d}{dy} F_{Y|i}(y)$$

$$f_Y(y) = \sum_i p(i) f_{Y|i}(y)$$

Consider a (n, N, λ) code $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ with $u_i \in \mathcal{X}^n$ and $A_i \subseteq \mathbb{R}^n$.

The error probability given that u_i is sent is

$$\lambda(u_i) = 1 - \int_{A_i} f_{Y|u_i(1)}(t_1) \cdots f_{Y|u_i(n)}(t_n) dt_1 \cdots dt_n.$$

Consider the partitioning of the real axis

$$\mathcal{P}: \quad \{b_1, b_2, \ldots, b_{M'-1}\}$$

which gives the partitions

$$\underbrace{(-\infty, b_1]}_{\mathcal{P}_1}, \quad \underbrace{(b_1, b_2]}_{\mathcal{P}_2}, \quad \ldots, \quad \underbrace{(b_{M'-1}, \infty)}_{\mathcal{P}_{M'}}$$

This can be seen as a discretization of the output Y. Let $[Y]_{\mathcal{P}}$ denote the discretized output. Since $[Y]_{\mathcal{P}}$ is discrete, it is clear that the channel between X and $[Y]_{\mathcal{P}}$ can be modeled as a DMC. The probability of the output Y taking the jth output value, i.e., belonging to partition $\mathcal{P}_j$ is defined to be

$$p(j) \triangleq \Pr\{Y \in \mathcal{P}_j\}.$$

The conditional probability of Y belonging to $\mathcal{P}_j$ given $X = a_i$ is given by

$$p(j|i) \triangleq \Pr\{Y \in \mathcal{P}_j | X = a_i\} = \int_{b_{j-1}}^{b_j} f_{Y|i}(t) dt.$$

Therefore,

$$p(j) = \Pr\{Y \in \mathcal{P}_j\} = \sum_{i=1}^{M} \Pr\{X = a_i\}\, p(j|i).$$

The mutual information of the effective DMC is denoted $I(X; [Y]_{\mathcal{P}})$. It can be shown that as the partitioning becomes finer, $I(X;[Y]_{\mathcal{P}})$ converge to

$$I(X;Y) = \sum_{i=1}^{M} p(i) \left(\int f_{Y|i}(t) \log_2 \frac{f_{Y|i}(t)}{f_Y(t)} dt \right). \tag{3}$$

where $f_Y(t) = \sum_i p(i) f_{Y|i}(t)$. We assume that the integral in (3) converge in the Riemann sense.

Theorem 4. *Let $\epsilon > 0$ and $0 < \lambda \leq 1$ be arbitrary. For n sufficiently large, there exists a code (n, N, λ) with $N \leq 2^{n(C-\epsilon)}$, where*

$$C \triangleq \max_{p(i)} \sum_i p(i) \int f_{Y|i}(y|i) \log_2 \frac{f_{Y|i}(y|i)}{f_Y(y)} dy.$$

For the DMC we have

$$N^{\text{dmc}} \geq 2^{n\left\{ I(X;[Y]_{\mathcal{P}}) - \frac{K_7(M,M',\lambda,K_0)}{\sqrt{n}} \right\}}.$$

which we obtained from [2, eq. (22)]. As the partitioning becomes finer, $I(X;[Y]_{\mathcal{P}})$ approaches C, but there will still be some finite gap. Therefore, for any given $\epsilon > 0$, we can always find some partitioning $\mathcal{P}$ such that

$$I(X;[Y]_{\mathcal{P}}) = C - \epsilon/2.$$

So, the achievable rate is

$$R = C - \epsilon/2 - K_7/\sqrt{n}.$$

Hence, for n sufficiently large, $R = C - \epsilon$ is achievable.

III. Entropy Typicality for Memoryless Continuous-Valued Random Variables

Definition 1. For a constant $\epsilon > 0$ and any n, we define the typical set $\mathcal{A}_\epsilon^{(n)}$ with respect to $f(y)$ as

$$\mathcal{A}_\epsilon^{(n)} = \left\{ (y_1, \ldots, y_n) \in \mathbb{R}^n \middle| \left| -\frac{1}{n} \log f(y_1, \ldots, y_n) - h(y) \right| \leq \epsilon \right\},$$

where

$$h(y) \triangleq -\int f_Y(y) \log_2 f_Y(y) dy$$

is the differential entropy.

Theorem 5. *(AEP) Let $Y_1, \ldots, Y_n$ be a sequence of i.i.d. random variables with distribution $f_Y(y)$. Then*

$$-\frac{1}{n} \log f(Y_1, \ldots, Y_n) \to \mathrm{E}_Y \left\{ \log \frac{1}{f(Y)} \right\}$$

in probability as $n \to \infty$.

Definition 2. The volume $\mathrm{Vol}(\mathcal{A})$ of a set $\mathcal{A} \subset \mathbb{R}^n$ is defined as

$$\mathrm{Vol}(\mathcal{A}) = \int_{x^n \in \mathcal{A}} dx_1 \cdots dx_n$$

Theorem 6. *The typical set $\mathcal{A}_\epsilon^{(n)}$ has the following properties:*

1) $\Pr\left\{ Y^n \in \mathcal{A}_\epsilon^{(n)} \right\} > 1 - \epsilon$ *for sufficiently large n.*
2) $\mathrm{Vol}\left(\mathcal{A}_\epsilon^{(n)} \right) \leq 2^{n(h(Y)+\epsilon)}$ *for all n.*
3) $\mathrm{Vol}\left(\mathcal{A}_\epsilon^{(n)} \right) \geq (1-\epsilon) 2^{n(h(Y)-\epsilon)}$ *for sufficiently large n.*

Proof: We prove 2) and 3). From Def. 1, for any $y^n \in \mathcal{A}_\epsilon^{(n)}$ we have

$$2^{-n(h(Y)+\epsilon)} \leq f_{Y^n}(y^n) \leq 2^{-n(h(Y)-\epsilon)}. \tag{4}$$

2) We know that

$$\int_{\mathcal{A}_\epsilon^{(n)}} f_{Y^n}(y^n) dy^n < 1.$$

Then, from (4) it follows that

$$1 > \int_{\mathcal{A}_\epsilon^{(n)}} f_{Y^n}(y^n) dy^n \geq 2^{-n(h(Y)+\epsilon)} \int_{\mathcal{A}_\epsilon^{(n)}} dy^n = 2^{-n(h(Y)+\epsilon)} \operatorname{Vol}\big(\mathcal{A}_\epsilon^{(n)}\big)$$
$$\Rightarrow \quad \operatorname{Vol}\big(\mathcal{A}_\epsilon^{(n)}\big) \leq 2^{n(h(Y)+\epsilon)}.$$

3) From 1) and (4) it follows that

$$1-\epsilon < \Pr\big\{Y^n \in \mathcal{A}_\epsilon^{(n)}\big\} = \int_{\mathcal{A}_\epsilon^{(n)}} f_{Y^n}(y^n) dy^n \leq 2^{-n(h(Y)-\epsilon)} \int_{\mathcal{A}_\epsilon^{(n)}} dy^n \leq 2^{-n(h(Y)-\epsilon)} vol \mathcal{A}_\epsilon^{(n)}$$
$$\Rightarrow \quad \operatorname{Vol}\big(\mathcal{A}_\epsilon^{(n)}\big) \geq (1-\epsilon) 2^{n(h(Y)-\epsilon)}.$$

■

REFERENCES

[1] A. Pitarokoilis, “Information Theory for Wireless Communications: Lecture 4.”
[2] H. Q. Ngo, “Information Theory for Wireless Communications: Lecture 3: Conditional Typicality, Channel Coding Theorem for DMC.”