Multi-use Information Theory 2 — Nov 1st, 2010

Lecture 4

Lecturer: Haim Permuter — *Scribe: Avihay Shirazi*

I. Introduction

This lecture contain two parts. In Section I we provide some properties of joint and conditional typical sets and in Section II we provide the achievability proof for the rate distortion theorem, using the results we developed in Section I.

II. Strong Typical Set

A reminder:

$$P_{x^n}(a) = \frac{N(a|x^n)}{n} \tag{1}$$

$$T_\epsilon^{(n)}(P_X) = T_\epsilon^{(n)}(X) = \left\{x^n \in \mathcal{X}^n : \ |P_{x^n}(a) - P_X(a)| \le \epsilon P_X(a)\right\} \tag{2}$$

$$|T(P)| \doteq 2^{nH(P)} \tag{3}$$

$$x^n \in T(P), \ \ Q^n(x^n) = 2^{-n\left(H(P)+D(P||Q)\right)} \tag{4}$$

$$\tag{5}$$

A. Jointly Strong Typical Set

Let us define

$$T_\epsilon^{(n)}(P_{XY}) = T_\epsilon^{(n)}(X,Y) = \left\{x^n, y^n : \ |P_{x^n y^n}(a,b) - P_{XY}(a,b)| \le \epsilon P_{XY}(a,b)\right\}. \tag{6}$$

Lemma 1 Let $x^n \in T_\epsilon^{(n)}(P_X)$ and $Y^n \sim$ i.i.d. P_Y then

$$\Pr\left\{(x^n, Y^n) \in T_\epsilon^{(n)}(P_{XY})\right\} \le 2^{-n\left(I(X;Y) - \delta(\epsilon)\right)}, \tag{7}$$

where $\delta(\epsilon, n) = \delta_{\epsilon,n} \to 0$ as $\epsilon \to 0$ and $n \to \infty$.

Properties:

1) If $X^n \sim$ i.i.d. P_X, then

$$\lim_{n\to\infty} \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} = 1 \tag{8}$$

or equivalently

$$1-\delta(\epsilon) \leq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} \leq 1, \tag{9}$$

where $\forall\epsilon > 0,\ \lim_{n\to\infty} \delta_{\epsilon,n} = 0$.

Proof: Because of the L.L.N. ■

2) If $X^n \in T_\epsilon^{(n)}(X)$, then

$$2^{-nH(X)(1+\epsilon)} \overset{(i)}{\leq} p(x^n) = \prod_{i=1}^{n} P_X(x_i) \overset{(ii)}{\leq} 2^{-nH(X)(1-\epsilon)}. \tag{10}$$

Proof: notice that $p(x^n) = \prod_{a\in\mathcal{X}} P_X(a)^{N(a|x^n)}$, hence, the left hand side of the inequality (i) follows from

$$\begin{aligned}
\frac{1}{n}\log p(x^n) &= \frac{1}{n}\sum_{a\in\mathcal{X}} N(a|x^n)\log P_X(a) && (11)\\
&= \sum_{a\in\mathcal{X}} P_{x^n}(a)\log P_X(a) && (12)\\
&\overset{(a)}{\geq} \sum_{a\in\mathcal{X}} (1+\epsilon)P_X(a)\log P_X(a) && (13)\\
&= -H(X)(1+\epsilon), && (14)
\end{aligned}$$

where step (a) follows from (2).

The right hand side of the inequality (ii) follows from

$$\begin{aligned}
\frac{1}{n}\log p(x^n) &= \frac{1}{n}\sum_{a\in\mathcal{X}} N(a|x^n)\log P_X(a) && (15)\\
&= \sum_{a\in\mathcal{X}} P_{x^n}(a)\log P_X(a) && (16)\\
&\leq \sum_{a\in\mathcal{X}} (1-\epsilon)P_X(a)\log P_X(a) && (17)\\
&= -H(X)(1-\epsilon). && (18)
\end{aligned}$$

■

3) For $X^n \sim$ i.i.d. P_X,

$$(1-\delta_{\epsilon,n})2^{nH(X)(1-\epsilon)} \overset{(i)}{\leq} |T_\epsilon^{(n)}(X)| \overset{(ii)}{\leq} 2^{nH(X)(1+\epsilon)}. \tag{19}$$

Proof: recall that $(1-\delta) \leq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} \leq 1$, then, the left hand side inequality (i)

follows from

$$
\begin{aligned}
\Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} &= \Pr\left\{\bigcup_{x^n \in \mathcal{X}^n} \left[\left(X^n = x^n\right) \cap \left(X^n \in T_\epsilon^{(n)}(X)\right)\right]\right\} && (20)\\
&\overset{(a)}{=} \sum_{x^n \in \mathcal{X}^n} \Pr\left\{\left(X^n = x^n\right) \cap \left(X^n \in T_\epsilon^{(n)}(X)\right)\right\} && (21)\\
&= \sum_{x^n \in \mathcal{X}^n} p(x^n) \cdot \mathbf{1}_{x^n \in T_\epsilon^{(n)}(X)} && (22)\\
& \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n) && (23)\\
&\leq \sum_{x^n \in T_\epsilon^{(n)}(X)} 2^{-n(H(X)(1-\epsilon)} && (24)\\
&= |T_\epsilon^{(n)}(X)| 2^{-nH(X)(1-\epsilon)}, && (25)
\end{aligned}
$$

and because $(1-\delta) \leq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\}$, we get that $(1-\delta)2^{nH(X)(1-\epsilon)} \leq |T_\epsilon^{(n)}(X)|$. Step (a) follows from the fact that the groups $\{X^n = x'^n\}$ and $\{X_n = x''^n\}$ are disjoint for $x'^n \neq x''^n$. i.e., if $X^n = x'^n$, then necessarily $X^n \neq x''^n$.

The right hand side inequality (ii) follows from

$$
\begin{aligned}
1 &\geq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} && (26)\\
&= \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n) && (27)\\
&\geq \sum_{x^n \in T_\epsilon^{(n)}(X)} 2^{-nH(X)(1+\epsilon)} && (28)\\
&= |T_\epsilon^{(n)}(X)| 2^{-nH(X)(1+\epsilon)}, && (29)
\end{aligned}
$$

therefore, $|T_\epsilon^{(n)}(X)| \leq 2^{nH(X)(1+\epsilon)}$. ■

B. *Conditionally strong typical set*

For a given $x^n \in \mathcal{X}^n$, let us define

$$T_\epsilon^{(n)}(P_{XY}|x^n) = \left\{y^n : \ (x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\right\}. \tag{30}$$

Notice that if $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$, then surely $x^n \in T_\epsilon^{(n)}(X)$.

Properties:

1) If $x^n \in T_{\epsilon_x}^{(n)}(X),\ p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ (DMC), $\epsilon_x \leq \epsilon$, then

$$1 - \delta_{\epsilon,\epsilon_x,n} \leq \Pr\left\{y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)\right\} \leq 1 \tag{31}$$

where $\forall \epsilon, \epsilon_x > 0,\ \delta_{\epsilon,\epsilon_x,n} \to 0$ as $n \to \infty$.

Proof: Follows directly from the L.L.N. ■

2) If $y^n \in T_\epsilon^{(n)}(P_{XY}|x^n),\quad p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ (DMC), then

$$2^{-nH(Y|X)(1+\epsilon)} \overset{(i)}{\leq} p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i) \overset{(ii)}{\leq} 2^{-nH(Y|X)(1-\epsilon)}. \tag{32}$$

Proof: Notice that

$$p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i) = \prod_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{Y|X}(b|a)^{N(a,b|x^n,y^n)}. \tag{33}$$

Now, the left hand side of the inequality (i) follows from

$$\begin{aligned}
\frac{1}{n}p(y^n|x^n) &= \frac{1}{n}\sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} N(a,b|x^n,y^n)\log P_{Y|X}(b|a) &(34)\\
&= \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{x^n,y^n}(a,b)\log P_{Y|X}(b|a) &(35)\\
&\geq \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} (1+\epsilon)P_{XY}(a,b)\log P_{Y|X}(b|a) &(36)\\
&= -H(Y|X)(1+\epsilon). &(37)
\end{aligned}$$

The right hand side of the inequality (ii) follows from

$$\begin{aligned}
\frac{1}{n}p(y^n|x^n) &= \frac{1}{n}\sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} N(a,b|x^n,y^n)\log P_{Y|X}(b|a) &(38)\\
&= \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{x^n,y^n}(a,b)\log P_{Y|X}(b|a) &(39)\\
&\leq \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} (1-\epsilon)P_{XY}(a,b)\log P_{Y|X}(b|a) &(40)\\
&= -H(Y|X)(1-\epsilon). &(41)
\end{aligned}$$

■

3) Given $x^n \in T_{\epsilon_x}^{(n)}(X)$, $p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ (DMC), $\epsilon_x \leq \epsilon$, then

$$(1-\delta_{\epsilon,\epsilon_x,n})2^{nH(Y|X)(1+\epsilon)} \overset{(i)}{\leq} |T_\epsilon^{(n)}(P_{XY}|x^n)| \overset{(ii)}{\leq} 2^{nH(Y|X)(1-\epsilon)}. \tag{42}$$

The proof is done in a similar way to the proof of (19).

Lemma 2 Consider a joint PMF P_{XY}. Let $x^n \in T_{\epsilon_x}^{(n)}(X)$ and y^n drawn i.i.d. according to P_Y and independent of x^n, then

$$2^{-n\left(I(X;Y)+\delta_\epsilon\right)} \overset{(i)}{\leq} \Pr\left\{Y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)\right\} \overset{(ii)}{\leq} 2^{-n\left(I(X;Y)-\delta_\epsilon\right)} \tag{43}$$

for $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

Proof: The left hand side inequality (i) follows from

$$\begin{aligned}
\Pr\left\{Y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)\right\} &= \sum_{y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)} p(y^n) && (44)\\
&\geq \sum_{y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)} 2^{-nH(Y)(1+\epsilon)} && (45)\\
&= |T_\epsilon^{(n)}(P_{XY}|x^n)\}|2^{-nH(Y)(1+\epsilon)} && (46)\\
&\geq (1-\delta_{\epsilon,n})2^{n(H(Y|X)(1-\epsilon)}2^{-nH(Y)(1+\epsilon)} && (47)\\
&= (1-\delta_{\epsilon,n})2^{-n\left(I(X;Y)+\delta_{\epsilon,n}\right.}. && (48)
\end{aligned}$$

The right hand side inequality (ii) follows from

$$\begin{aligned}
\Pr\left\{Y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)\right\} &= \sum_{y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)} p(y^n) && (49)\\
&\leq \sum_{y^n \in T_\epsilon^{(n)}(P_{XY}|x^n)} 2^{-nH(Y)(1-\epsilon)} && (50)\\
&= |T_\epsilon^{(n)}(P_{XY}|x^n)|2^{-nH(X)(1-\epsilon)} && (51)\\
&\leq 2^{-n\left(H(Y)-H(Y|X)\right)+n\epsilon\left(H(Y)+H(Y|X)\right)} && (52)\\
&= 2^{-n\left(I(X;Y)-\delta_\epsilon\right)}. && (53)
\end{aligned}$$

■

III. Achievability of the Rate-Distortion theorem

The rate distortion function for the source $X \sim$ i.i.d. $P_X(x)$ and the distortion measure D is given by

$$R(D) = \min_{p(\hat{x}|x):E[d(\hat{X},X)]\leq D} I(X;\hat{X}). \tag{54}$$

We proved the converse in the previous lecture, let us now prove the achievability. The following lemma will be used in the error analysis part of the achievability proof.

Lemma 3 For $0 \leq X \leq 1$ and for any $n \in \mathbb{N}$, the following inequality holds

$$(1-x)^n \leq e^{-nx}. \tag{55}$$

Proof: Let $f(x) = e^{-x} - (1-x)$, then, $f'(x) = 1 - e^{-x} > 0$ for $x > 0$ and hence $f(x) > 0$ for $x > 0$. Therefore, $1 - x \leq e^{-x}$ for $0 \leq x \leq 1$, and by raising both sides to the power of n, we get that $(1-x)^n \leq e^{-nx}$. ■

Sketch of the Achievability: We are given by nature the source $X \sim$ i.i.d. P_X. For the distortion measure D, we consider the mappings $f: \ \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$ and $g: \ \{1, 2, \ldots, 2^{nR}\} \to \hat{\mathcal{X}}^n$ such that $\mathbb{E}\big[d(X^n, \hat{X}^n)\big] \leq D$, where $d(X^n, \hat{X}^n) = \sum_{a,b} P_{x^n, \hat{x}^n}(a,b) d(a,b)$. We will design a code such that $(X^n, \hat{X}^n) \in T_\epsilon^{(n)}(P_{X\hat{X}})$ with probability 1. Then, we will show that if $R > I(X; \hat{X})$, such a code does exist.

Achievability

1) *Code design:* Generate at random 2^{nR} codewords $\hat{X}^n \sim$ i.i.d. $P_{\hat{X}}$, and index them $\big\{\hat{X}^n(1), \hat{X}^n(2), \ldots, \hat{X}^n(2^{\lceil nR \rceil})\big\}$.

2) *Encoder:* Given the word x^n, look for a codeword $\hat{X}^n(k)$ s.t. $\big(x^n, \hat{X}^n(k)\big) \in T_\epsilon^{(n)}(X, \hat{X})$, and send the index k. If such a codeword does not exist, send 0.

3) *Decoder:* Given the index k, declare $\hat{X}^n = \hat{X}^n(k)$.

Analysis of an error: For a given $x^n \in \mathcal{X}$, define the events

$$\begin{aligned} E_1 &:= \big\{x^n \notin T_\epsilon^{(n)}(X)\big\}, \\ E_2 &:= \big\{\hat{X}^n(k) \notin T_\epsilon^{(n)}(P_{X\hat{X}}|x^n) \ \ \forall k \in \{1, 2, \ldots, 2^{nR}\}\big\}. \end{aligned}$$

From union of bounds, we know that $P_e^{(n)} \leq \Pr\{E_1\} + \Pr\{E_2\}$. Now, $\Pr\{E_1\} \to 0$ because of the L.L.N., as for E_2,

$$\Pr\{E_2\} = \Pr\big\{\nexists k \in \{1, 2, \ldots, 2^{nR}\}: \ \ X^n(k) \in T_\epsilon^{(n)}(P_{X\hat{X}}|x^n)\big\} \tag{56}$$

$$= \prod_{i=1}^{2^{nR}} \Pr\big\{\hat{X}^n(1) \notin T_\epsilon^{(n)}(P_{X\hat{X}}|x^n)\big\} \tag{57}$$

$$\leq \prod_{i=1}^{2^{nR}} (1 - 2^{-n\left(I(X;\hat{X}) - \delta_\epsilon\right)}) \tag{58}$$

$$\stackrel{(a)}{\leq} \exp\{-2^{nR}2^{-nI(X;\hat{X})+n\delta_\epsilon}\} \tag{59}$$

$$= \exp\{-2^{n\left(R-I(X;\hat{X})+n\delta_\epsilon\right)}\}, \tag{60}$$

where the last term tends to 0 for $R > I(X;\hat{X})$, and this completes the proof of the rate-distortion theorem. Step (a) follows from lemma 3.