Mathematical methods in communication June 23th, 2009

Lecture 11

Lecturer: Haim Permuter *Scribe: Amit Toren and Liran Mishali*

I. Strong Typicality Set

We define Weak Typicality set as: *(Weak typicality)*

$$A_{\epsilon}^{(n)} = \left\{ x^n \in \mathcal{X}^n : \left| -\frac{1}{n} log P(x^n) - H\left(X\right) \right| \leq \epsilon \right\}. \tag{1}$$

The expression $N(a|x^n) =$ is defined as the number of appearances of symbol a in the sequence x^n
Example: $x^n =$ 01011110 => $N(0|x^n) = 3$, $N(1|x^n) = 5$

Definition 1 (Strong Typicality) A sequence $x^n \in \S^n$ is said to be $\epsilon - strongly\ typical$ with respect to a distribution $P(x)$ on $\mathcal{X}$ if:

- For all a $\in \mathcal{X}$ with $P_X(a) > 0$, we have:

$$\left| \frac{N(a|x^n)}{n} - P_X(a) \right| \leq \frac{\epsilon}{|\mathcal{X}|} \tag{2}$$

- For all $a \in \mathcal{X}$ with $P_X(a) = 0$, $N(a|x^n) = 0$.

Lemma 1 For $X \sim i.i.d.$ and the expression : $\frac{N(a|x^n)}{n}$ if we take n $\to \infty$ then we get:

$$\frac{N(a|x^n)}{n} \to P_X(a)$$

$Proof$:

$$N(a|x^n) = \sum_{i=1}^{n} 1_a(x_i) \tag{3}$$

$$1_a(X_i) = \begin{cases} 1 & X_i = a \\ 0 & X_i \neq a \end{cases} \tag{4}$$

By the Law of large numbers , for any $\delta \geq 0$, $\epsilon > 0$ $\exists n$ s.t

$$\Pr\left(\left| \frac{N(a|x^n)}{n} - P_X(a) \right| < \epsilon \right) \geq 1 - \delta$$

Theorem 1 The typical set T_{ϵ}^{n} has the following properties

1) If $x^n \in T_{\epsilon}^{(n)}(x)$ then:

$$H(X) - \epsilon_1 \leq -\frac{1}{n} log P(x^n) \leq H(X) + \epsilon_1 \tag{5}$$

2) For all $\delta \geq\ 0$ exists n sufficiently large s.t $\Pr(x^n \in T_\epsilon^{(n)}(x)) \geq 1-\delta$

3) $2^{n(H(x)-\epsilon_2)} \leq \left|T_\epsilon^{(n)}(x)\right| \leq 2^{n(H(x)+\epsilon_2)}$

$Proof$ (1):

$$\begin{aligned} -\frac{1}{n}logP_X(x^n) &\overset{X \backsim i.i.d}{=} -\frac{1}{n}log\prod_{i=1}^{n}P_X(x^n) \\ &= -\frac{1}{n}\Sigma_{i=1}^{n}logP_X(x^n) \\ &= -\frac{1}{n}\Sigma_{a\in\mathcal{X}}N(a|x^n)logP_X(x^n) \end{aligned}$$

Example 1 For the series $x^n = 0001011$ with probabilities: $P(0)=\frac{1}{4}, P(1)=\frac{3}{4}$

$$N(0|x^n)=4, N(1|x^n)=3$$

Instead of summing $log\frac{1}{4}+log\frac{1}{4}+log\frac{1}{4}+log\frac{3}{4}+log\frac{1}{4}$......

We will multiply the number of zeroes and ones in the the corresponded entropy

$$\begin{aligned} N(0|x^n)log\frac{1}{4}+N(1|x^n)log\frac{3}{4} &= \left|-H(X)-\frac{1}{n}logP_X(x^n)\right| \\ &= \left|\Sigma_{a\in\mathcal{X}}P_X(a)logP_X(a)-\frac{1}{n}logP_X(x^n)\right| \\ &= \left|\Sigma_{a\in\mathcal{X}}(P_X(a)-\frac{N(a|x^n)}{n})-logP_X(a)\right| \\ &\leq \frac{\epsilon}{|X|}\Sigma_{a\in\mathcal{X}}\left|logP_X(a)\right| = \epsilon_1 \end{aligned}$$

Explanation of (3):

Lets assume that our series is a series with the length of N

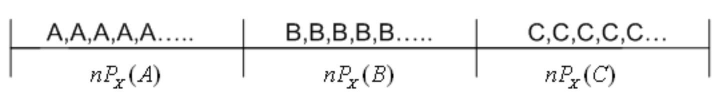

Fig. 1. All possible sequences

$$\begin{aligned} \left(\frac{n!}{nP_x(a)!nP_x(b)!nP_x(c)!)}\right) &= \frac{logn!}{nlogn} \\ &= \frac{nlogn-n+\frac{1}{2}log2\Pi n}{nlogn} \overset{n\to\infty}{\to} 1 \\ \# = & \frac{n^n}{(nP_x(a))^{nP(a)}(nP_x(b))^{nP(b)}(nP_x(c))^{nP(c)}} = K \end{aligned}$$

K - Number of sequences

$logK = -nP_x(a)\log P_x(a) - nP_x(b)\log P_x(b) - nP_x(c)\log P_x(c) = nH(X)$

Definition 2 (Joint Typical Set)

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

If $P_{X,Y}(a,b) = 0, N(a,b|x^n,y^n) = 0$

Definition 3 (Conditional strongly typical set)
Let $y^n \in T_\epsilon^{(n)}(Y)$ then:

$$T(X|y^n) = \{x^n : (x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\} \tag{7}$$

$$|T(X|y^n)| = 2^{nH(X|Y)}$$

$$T(Y|x^n) = \{y^n : (y^n, x^n) \in T_\epsilon^{(n)}(X,Y)\} \tag{8}$$

$$|T(Y|x^n)| = 2^{nH(Y|X)}$$

$$|T_\epsilon{}^n(X,Y)| = 2^{nH(Y,X)} \; |T_\epsilon{}^n(X|Y)| = 2^{nH(X|Y)} \tag{9}$$

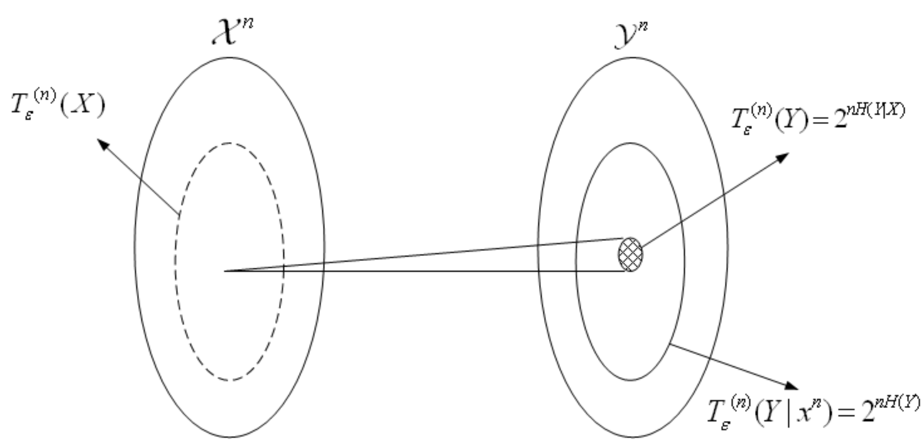

Fig. 2. Noisy Typewriter From X to Y

Example 2 We will show that $|T_\epsilon{}^n(Y|X)| = 2^{nH(Y|X)}$

The channel transfer a to d/e according to the channel noise , b to f/g etc.

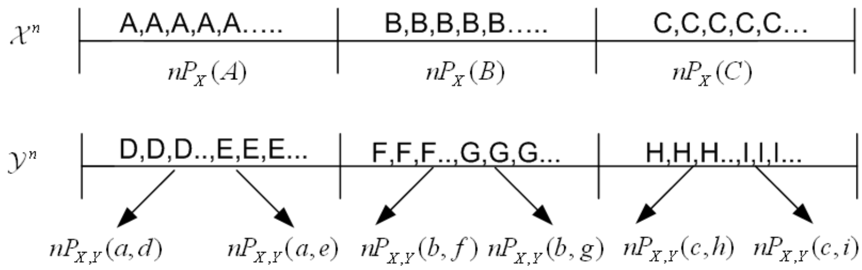

Fig. 3. Example 2

The amount of series for all inputs :

$$|T_\epsilon^n(Y|X)| = \binom{nP_X(a)}{nP_{X,Y}(a,d)nP_{X,Y}(a,e)}\binom{nP_X(b)}{nP_{X,Y}(b,f)nP_{X,Y}(b,g)}\binom{nP_X(c)}{nP_{X,Y}(c,h)nP_{X,Y}(c,i)}$$

Lets use the approximation : $n! \approx n^n$ and operate log, and each binom will be:

$-H(X) + H(Y,X) = H(Y|X)$

Applying it to the whole expression we will get: $|T_\epsilon{}^n(Y|X)| = 2^{nH(Y|X)}$ ■

II. Rate distortion

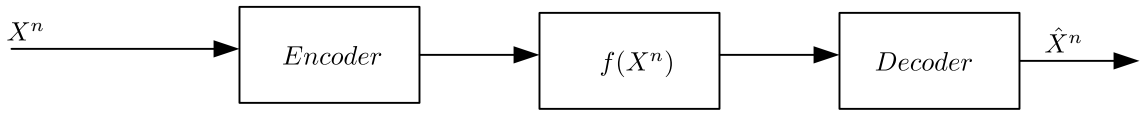

Fig. 4. Communication system

Definition 4 (Distortion function) A distortion function or distortion measure is a mapping

$$d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^+ \tag{10}$$

from the set of source alphabet-reproduction alphabet pairs into the set of nonnegative real numbers. The distortion $d(x, \hat{x})$ is a measure of the cost of representing the symbol x by the symbol $\hat{x}$.

Definition 5 (Distortion Bound) A distortion measure is said to be bounded if the maximum value of the distortion is finite:

$$d_{max} \stackrel{def}{=} \max_{x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x}) < \infty. \tag{11}$$

In most cases, the reproduction alphabet $\hat{\mathcal{X}}$ is the same as the source alphabet $\mathcal{X}$.

Example 3 : Examples of common distortion function are:

$$d(X_i, \hat{X}_i) = X_i \oplus \hat{X}_i - Hamming\ Distance$$

$$d(X_i, \hat{X}_i) = (X_i - \hat{X}_i)^2 - Mean\ Square\ Error$$

Definition 6 (Distortion between sequences) The distortion between sequences x^n and $\hat{x^n}$ is defined by:

$$D(X^n, \hat{X^n}) = \frac{1}{n} \Sigma_{i=1}^{n} d(X_i, \hat{(}X_i)) \tag{12}$$

So the distortion for a sequence is the average of the per symbol distortion of the elements of the sequence.

Definition 7 ($(2^{nR}, n)$-rate distortion code)
A $(2^{nR}, n)$-rate distortion code consists of:
Encoder:$f(X^n) : X^n \rightarrowtail (1, 2, 3, , , 2^{nR})$
Decoder:$g(f(X^n)) : X^n \rightarrowtail (1, 2, 3, , , 2^{nR})$
The distortion associated with the $(2^{nR}, n)$ code is defined as $\bar{D}(X^n, \hat{X^n}) = \frac{1}{n} \Sigma_{i=1}^{n} d(X_i, \hat{X}_i)$

Definition 8 (Achivable Rate)

A rate distortion pair (R, D) is achivable if $\exists$ a sequence of $(n, 2^{nR})$ codes s.t : $\lim_{n\to\inf} \bar{D}(X^n, \hat{X^n}) \leq D$
$R(D)^{(I)} = \min_{P(\hat{x}|x):E(d(x,\hat{x}))\leq D} I(X;\hat{X})$

Where the minimization is over all conditional distributions $P(\hat{x}|x)$ for which the joint distribution $P(x|\hat{x}) = P(x)P(\hat{x}|x)$ satisfies the expected distortion constrained.

Definition 9 (Rate Distortion lower bound)

The rate distortion function $R(D)$ is the infimum of all R that are achievable with Distortion D

Definition 10 (Distortion Rate lower bound)

The distortion rate function $D(R)$ is the infimum of all distortion D such that (R, D) is in the rate distortion region of the source for a given rate R.

Theorem 2 The rate distortion function for an i.i.d. source X with distribution p(x) and bounded distortion function $d(x, \hat{x})$ is equal to the associated rate distortion function. Thus,

$$R(D) = R(D)^{(I)} = \min_{P(\hat{x}|x):\sum_{(x,\hat{x})} P(x)P(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X}) \tag{13}$$

A. CALCULATION OF THE RATE DISTORTION FUNCTION

1) Binary Source:

Theorem 3 (The rate distortion function for a Bernoulli(p) source with Hamming distortion)
X∼Ber(p) ,$p \leq \frac{1}{2}, D \leq \frac{1}{2}$
$d(X_i, \hat{X}_i) = X_i \oplus \hat{X}$
$R(D) =?$
$Proof$

$$R(D) = \begin{cases} H_b(p) - H(D) & p > D \\ 0 & p < D \end{cases}$$

If $D = 0$ $X_i = \hat{X}_i \Rightarrow R = H_b(p)$

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(X) - H(X \oplus \hat{X}_i|\hat{X}) \\ &\geq H(X) - H(X \oplus \hat{X}_i) \\ & H_b(p) - H_b(D) \end{aligned}$$

$$We\ demand\ : E[d(X_i,\hat{X}_i)] \leq D\ \ ,\ P_r[X_i \oplus \hat{X}_i = 1] \leq D$$

We will achive it with:

$$X \backsim Ber(p),\ X = \hat{X} \oplus Z,\ Z \backsim Ber(p),\ \ Z \perp \hat{X}$$

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(X) - H(X \oplus \hat{X}_i|\hat{X}) \\ &\geq H(X) - H(Z) \\ & \quad H_b(p) - H_b(D) \end{aligned}$$

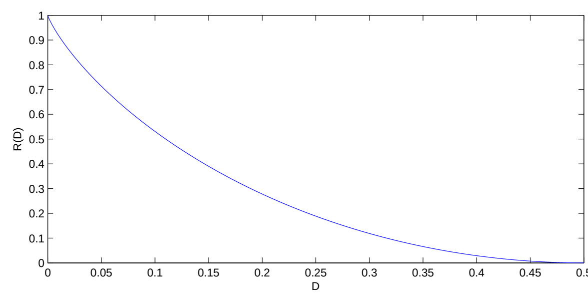

Fig. 5. Rate distortion function for a Bernoulli ($\frac{1}{2}$) source.

Theorem 4 (The rate distortion function for a $\mathcal{N}(0,\sigma^2)$ source with squred-error distortion)

$$R(D) = \begin{cases} \frac{1}{2}log\frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2 \\ 0 & D \geq \sigma^2 \end{cases}$$

Proof : Let X be $\backsim \mathcal{N}(0,\sigma^2)$.By the rate distortion theorem extended to continuous alphabets, we have

$$R(D) = \min_{f(\hat{x}|x):E(\hat{X}-X)^2 \leq D} I(X;\hat{X}). \tag{14}$$

First we should find the lower bound for the rate distortion function and prove that this is achievable.

$$\begin{aligned} I(X;\hat{X}) &= h(X) - h(X|\hat{X}) \\ &= \frac{1}{2}log(2\pi\epsilon)\sigma^2 - h(X-\hat{X}|\hat{X}) \\ &\geq \frac{1}{2}log(2\pi\epsilon)\sigma^2 - h(X-\hat{X}) \\ &\geq \frac{1}{2}log(2\pi\epsilon)\sigma^2 - h(\mathcal{N}(0,E(X-\hat{X})^2)) \\ &= \frac{1}{2}log(2\pi\epsilon)\sigma^2 - \frac{1}{2}log(2\pi\epsilon)E(X-\hat{X})^2 \end{aligned}$$

$$\geq \frac{1}{2}log(2\pi\epsilon)\sigma^2 - \frac{1}{2}log(2\pi\epsilon)D$$
$$= \frac{1}{2}log\frac{\sigma^2}{D}$$

Conclusion:

$$R(D) \geq \frac{1}{2}log\frac{\sigma^2}{D} \quad (15)$$

If $D \leq \sigma^2$ we choose

$$X = \hat{X} + Z, \hat{X} \backsim \mathcal{N}(0, \sigma^2 - D), Z \backsim \mathcal{N}(0, D)$$

where $\hat{X}$ and Z are independent. For this joint distribution, we calculate

$$I(X;\hat{X}) = \frac{1}{2}log\frac{\sigma^2}{D} \quad (16)$$

and $E(X-\hat{X})^2 = D$, thus achieving the bound.If $D > \sigma^2$, we choose $\hat{X} = 0$ with probability 1, achieving $R(D) = 0$. Hence, the rate distortion function for the Gaussian source with squared-error distortion is

$$R(D) = \begin{cases} \frac{1}{2}log\frac{\sigma^2}{D} & 0 \leq \text{D} \leq \sigma^2 \\ 0 & \text{D} \geq \sigma^2 \end{cases}$$

We can rewrite $R(D)$ as $D(R)$: $D(R) = \sigma^2 2^{-2R}$.

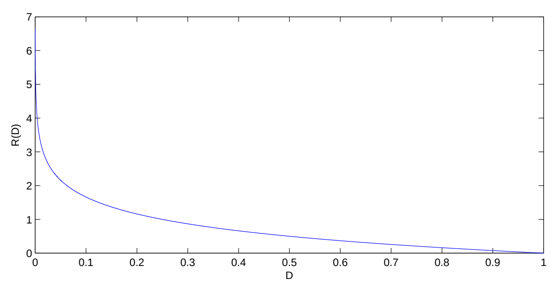

Fig. 6. Rate distortion function for a Gaussian source.