Mathematical methods in communication | Nov 24th, 2015

Lecture 2

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

I. Rate distortion

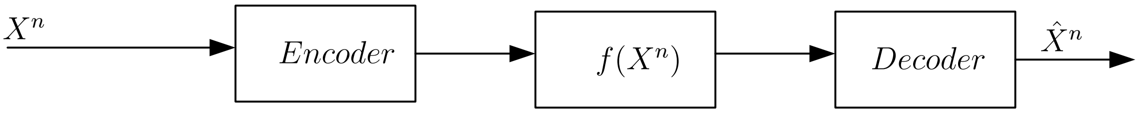

Fig. 1. Communication system

Definition 1 (Distortion function.) A distortion function or distortion measure is a mapping

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathcal{R}^+ \tag{1}$$

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

$$d_{max} \overset{def}{=} \max_{x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x}) < \infty. \tag{2}$$

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

Example 1 : 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 3 (Distortion between sequences) The distortion between sequences x^n and $\hat{x^n}$ is defined by:

$$D(x^n, \hat{x^n}) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, (\hat{x}_i)). \tag{3}$$

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

Definition 4 ($(2^{nR}, n)$-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: $g(f(X^n)) : (1, 2, 3, , , 2^{nR}) \rightarrowtail \hat{X}^n$

The distortion associated with the $(2^{nR}, n)$ code is defined as $E[D(X^n, \hat{X}^n)] = \frac{1}{n}\sum_{i=1}^{n} E[d(X_i, \hat{X}_i)]$

Definition 5 (Achivable Rate.) A rate distortion pair (R, D) is *achivable* if there exists a sequence of $(n, 2^{nR})$ codes s.t : $\lim_{n\to\infty} E[D(X^n, \hat{X}^n] \leq D$

Definition 6 (Rate Distortion function $R(D)$) Given a distortion D, *rate distortion function* $R(D)$ is the infimum of all achievable rates with Distortion D

Definition 7 (Distortion Rate function $D(R)$) Given a rate R, the *distortion rate function* $D(R)$ is the infimum of all distortion D such that (R, D) is in the rate distortion region.

Let us define the mathematical measure $R^{(I)}(D)$ as

$$R(D)^{(I)} = \min_{p(\hat{x}|x):E(d(x,\hat{x}))\leq D} I(X; \hat{X}), \tag{4}$$

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.

Theorem 1 The rate distortion function function an i.i.d. source $X \sim P_X$ and bounded distortion function $d(x, \hat{x})$ is equal to

$$R(D) = R^{(I)}(D) = \min_{p(x;\hat{x}):\sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X; \hat{X}) \tag{5}$$

A. CALCULATION OF THE RATE DISTORTION FUNCTION

1) Binary Source:

Example 2 Consider the rate distortion function for a Bernoulli(p) source with Hamming distortion. Mathematically, $X \sim \text{Ber}(p)$,$p \leq \frac{1}{2}, D \leq \frac{1}{2}$, $d(x_i, \hat{x}_i) = x_i \oplus \hat{x}_i$. What is the rate distorstion function, $R(D)$?

Solution

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

Note that if $D = 0$ then $X = \hat{X}$ and we obtain the well known result of lossless compression $R = H_b(p)$

Now, lets show that this is the best we can do (upper bound):

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

We will achive it with: $\hat{X} \sim Ber(p_0)$,$Z \sim Ber(D)$, $Z \perp \hat{X}$ and $X = \hat{X} \oplus Z$ where $X \sim Ber(p)$. The relation

$$p_0 * D \triangleq p_0(1-D) + (1-p_0)D = p \tag{6}$$

should hold. It implies $p_0 = \frac{p-D}{1-2D}$. Note that if $\frac{1}{2} \geq p > D$ such p_0 exists.

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

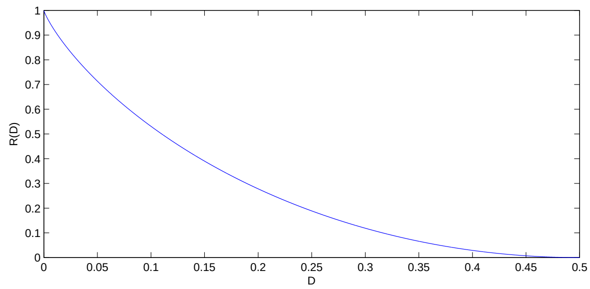

Fig. 2. Rate distortion function for a Bernoulli ($\frac{1}{2}$) source.

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

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

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

$$R(D) = \min_{f(\hat{x}|x):E(\hat{X}-X)^2 \le D} I(X;\hat{X}). \tag{7}$$

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

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

Conclution:

$$R(D) \geq \frac{1}{2}\log\frac{\sigma^2}{D} \tag{8}$$

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

$$I(X;\hat{X}) = \frac{1}{2}\log\frac{\sigma^2}{D} \tag{9}$$

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

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

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

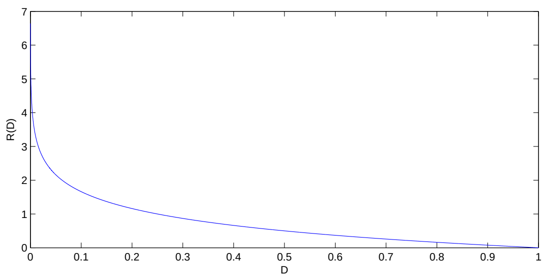

Fig. 3. Rate distortion function for a Gaussian source.

II. Strong Typicality Set

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

$$A_\epsilon^{(n)} = \left\{ x^n \in \mathcal{X}^n : \left| -\frac{1}{n} logP(x^n) - H\left(X\right) \right| \leq \epsilon \right\}. \tag{10}$$

The expresion $N(a|X^n) =$ 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 8 *(Strong Typicality)* A sequence $x^n \in \mathcal{X}^n$ 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:

$$\left|\frac{N(a|X^n)}{n} - P_X(a)\right| \leq \frac{\epsilon}{|\mathcal{X}|} \tag{11}$$

- 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 $\rightarrow \infty$ then we get:

$$\frac{N(a|X^n)}{n} \rightarrow P_X(a)$$

$Proof$:

$$N(a|X^n) = \sum_{i=1}^{n} 1_a(x_i) \tag{12}$$

$$1_a(X_i) = \begin{cases} 1 & X_i = a \\ 0 & X_i \neq a \end{cases} \tag{13}$$

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 3 The typical set A_ϵ^n has the following propertires

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

$$H(X) - \epsilon_1 \leq -\frac{1}{n} log P(x^n) \leq H(X) + \epsilon_1 \tag{14}$$

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):

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

Example 3 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

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

Definition 9 *(Joint Typical Set)*

$$T_\epsilon^{(n)}(X,Y) = \{X^n, Y^n : \left|\frac{N(a,b|X^n,Y^n)}{n} - P_xy(a,b)\right| \leq \frac{\epsilon}{|X||Y|}\} \tag{15}$$

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

Definition 10 *(Conditional strongly typical set)*

Let $Y^n \in T_\epsilon^{(n)}(Y)$ then:

$$T(X|Y^n) = \{X^n : (X^n,Y^n) \in T_\epsilon^{(n)}(X,Y)\} \tag{16}$$

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

$$T(Y|X^n) = \{Y^n : (Y^n,X^n) \in T_\epsilon^{(n)}(X,Y)\} \tag{17}$$

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

$$|T_\epsilon{}^n(X,Y)| = 2^{nH(Y,X)}\ |T_\epsilon{}^n(X|Y)| = 2^{nH(X|Y)} \tag{18}$$