| 6.441 Transmission of Information | Feb. 16, 2006 |
|---|---|
| Lecture 4 | |
| *Lecturer:* Madhu Sudan | *Scribe:* Costas Pelekanakis (gas) |

# 1. Today's outline

a. Asymptotic Equipartition Property (A.E.P.)
b. Typical sets
c. Application to data compression

# 2. Review of last lecture

Notions:

- $H(x), H(x/y), I(x;y), I(x;y/z)$
- $D(p//q) = \sum_x p(x)\log\frac{p(x)}{q(x)}$ which is measure of the inefficiency of assuming that the distribution of x is q(x) when the true distribution is p(x).
- Markov chain: $X \to Y \to Z$ or equivalently $p_{x,y/z} = p_{x/y}p_{z/y}$

Results:

- $D(p//q) \geq 0$ with equality when $p(x) = q(x)$
- $I(x;y) = D(p_{x,y} // p_x p_y) \geq 0$
- $H(x/y) = H(x) - I(x;y) \leq H(x)$
- chain rule: $H(x_1,...,x_n) \leq \sum_{i=1}^{n} H(x_i)$
- If $X \to Y \to Z$ then $I(x;z) \leq I(x;y)$
- If $X \to Y \to \hat{X}$ then Fano's inequality: $\Pr(X \neq \hat{X}) = P_e \geq \frac{H(x/y)-1}{\log|\Omega_x|}$

## 2.1. Review of Fano's inequality

We give two examples which show that Fano's inequality can be either weak or tight.

### 2.1.1. Example 1

*x is uniformly distributed over the set of binary n-tuples and y takes values from the set of binary n/2-tuples. I claim no matter distribution I pick for y,* $H(x/y) \ge n/2$.

We have:

$$H(x) = \log|\Omega_x| = \log_2 2^n = n$$
$$H(y) \le \log|\Omega_y| = \log_2 2^{n/2} = n/2$$
$$H(x/y) = H(x) - I(x;y) = n - I(x;y)$$
$$I(x;y) = H(y) - H(y/x) \le H(y) \le n/2$$

Thus, $H(x/y) \ge n/2$.

Fano's inequality yields (assuming big n):

$$P_e \ge \frac{H(x/y)-1}{\log|\Omega_x|} \approx \frac{n/2}{n} = 0.5$$

A better bound on $P_e$ in this case is:

$$P(\text{correct decoding}) \le \frac{2^{n/2}}{2^n} \Rightarrow P(\text{error}) \ge 1 - \frac{2^{n/2}}{2^n}$$

In this example Fano's inequality is very weak!

### 2.1.2. Example 2

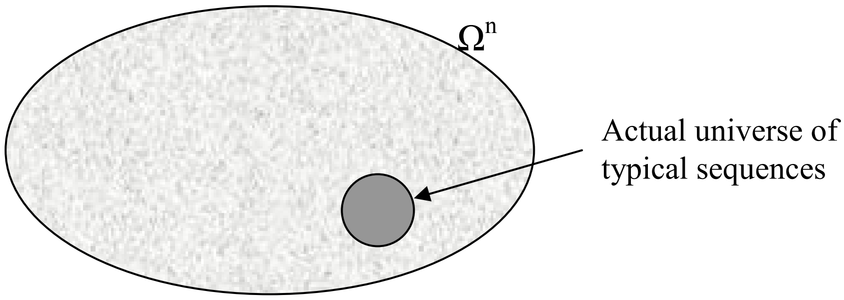
*x,y are distributed as follows: with probability p, x=y, and x,y are uniformly distributed over the set of binary m-tuples and with probability 1-p, x is uniformly distributed over the set of binary n-tuples and y is a constant.*

This is the picture of an erasure channel. The best strategy would decode as follows: observe y and assume that this is what it was sent. Obviously $P_e = p$ in this case. Fano's inequality yields:

$$H(x/y) = p \cdot H(x/y = const) + (1-p) \cdot H(x/y = x) = p \cdot n + 0 = p \cdot n$$
$$P_e \ge \frac{p \cdot n - 1}{n} \approx p$$

## 3. Typical sets

We want to answer the following question: if $x_1,\dots,x_n$ are iid and $x_i \sim p(x)$, i=1...n, what is the probability of the sequence $(x_1,\dots,x_n)$ to occur an n goes large? This will lead us to divide the set of the sequences into two sets, the typical set, which contains the “highly likely to occur” sequences and the non-typical set which contains all the other sequences.




We will use the law of large numbers to answer the above question.

### 3.1. A.E.P. Lemma

*If $x_1,\dots,x_n$ are i.i.d. according to p(x) then* $-\dfrac{\log p(x_1 \dots x_n)}{n} \to H(x)$ *in probability. In other words: for every ε>0, δ>0 there exists $n_o(\delta,\varepsilon)$ such that for every $n>n_o(\delta,\varepsilon)$ the* $\Pr\{H(x)-\varepsilon \le -\dfrac{\log p(x_1 \dots x_n)}{n} \le H(x)+\varepsilon\} \ge 1-\delta$. *(Actually δ goes to 0 as $\exp(-n\varepsilon^2)$).*

Proof: Note that $p(x_1,\dots,x_n)$ is the probability of observing the sequence $(x_1,\dots,x_n)$. We have that $x_i$ are iid so:

$$\frac{1}{n}\log p(x_1 \dots x_n) = \frac{1}{n}\log\prod_{i=1}^{n} p(x_i) = \frac{1}{n}\sum_{i=1}^{n}\log p(x_i)$$

Let us call a new r.v. $z_i = -\log p(x_i)$. The $z_i$’s are also i.i.d. and $E[z] = H(x)$. Applying the law of large numbers we have:

$$-\frac{1}{n}\sum_{i=1}^{n}\log p(x_i) = \frac{1}{n}\sum_{i=1}^{n} z_i \xrightarrow{LLN} E[z] = H(x)$$

Rewriting the above result we note that the probability to observe a sample sequence $(x_1,\dots,x_n)$ is bounded as: $2^{-n(H(x)+\varepsilon)} \le p(x_1,\dots,x_n) \le 2^{-n(H(x)-\varepsilon)}$. This motivates the definition of the typical set.

### 3.2. Definition: Typical set

$$A_\varepsilon^{(n)} = \{(x_1,...,x_n) \mid 2^{-n(H(x)+\varepsilon)} \le p(x_1,...,x_n) \le 2^{-n(H(x)-\varepsilon)}\}$$

### 3.3. Typical set theorem

i) $\Pr\{A_\varepsilon^{(n)}\} \ge 1-\delta$

ii) $\left|A_\varepsilon^{(n)}\right| \le 2^{n(H(x)+\varepsilon)}$

iii) $\left|A_\varepsilon^{(n)}\right| \ge (1-\delta)2^{n(H(x)-\varepsilon)}$

<u>Proof:</u>

i) A.E.P. Lemma

ii) $1 \ge \Pr\{A_\varepsilon^{(n)}\} \ge \left|A_\varepsilon^{(n)}\right| \cdot 2^{-n(H(x)+\varepsilon)} \Rightarrow \left|A_\varepsilon^{(n)}\right| \le 2^{n(H(x)+\varepsilon)}$

iii) $\left|A_\varepsilon^{(n)}\right| \cdot 2^{-n(H(x)-\varepsilon)} \ge \Pr\{A_\varepsilon^{(n)}\} \ge 1-\delta \Rightarrow \left|A_\varepsilon^{(n)}\right| \ge (1-\delta)\cdot 2^{n(H(x)-\varepsilon)}$

### 3.4. Example

Let $z = \begin{Bmatrix} 0 & \text{w.p. } 9/10 \\ 1 & \text{w.p. } 1/20 \\ -1 & \text{w.p. } 1/20 \end{Bmatrix}$

We expect the typical sequences $(z_1,...,z_n)$ to contain $\frac{9}{10}n(1\pm\varepsilon')$ "zeros", $\frac{1}{20}n(1\pm\varepsilon')$ "ones" and $\frac{1}{20}n(1\pm\varepsilon')$ "minus ones". Furthermore, $\left|A_\varepsilon^{(n)}\right| \approx 2^{n(H(z)\pm\varepsilon')}$

## 4. Application: Data compression

Compression is a mapping (function) of a higher dimensional space onto a lower one. Suppose we want to map the set $\Omega_x$ of binary n-tuples to the set $\Omega_y$ of the binary m-tuples with m«n. Obviously, the mapping is not "1-1" so errors will occur during decoding. We divide $\Omega_x$ into two sets: the $A_\varepsilon^{(n)}$ and its complement. We are computing the following quantity:

$$\Pr\{\text{decoding correctly}\} = \underbrace{\Pr\{\text{decoding correctly}/\Omega_x \setminus A_\varepsilon^{(n)}\}}_{\delta} + \Pr\{\text{decoding correctly}/A_\varepsilon^{(n)}\}$$

$$= \delta + \Pr\{(x_1,\ldots,x_n) \in A_\varepsilon^{(n)} \text{ and } (x_1,\ldots,x_n) \in \text{image of decoder}\}$$

$$\leq \delta + \left|\Omega_y\right| \max(\Pr\{(x_1,\ldots,x_n) \in A_\varepsilon^{(n)}\}) = \delta + 2^m \cdot 2^{-n(H(x)-\varepsilon)} \leq \delta + 2^m \frac{2^{2n\varepsilon}}{\left|A_\varepsilon^{(n)}\right|}$$