Lecture 17

Tuesday, November 18, 2014
8:59 AM

Ch. 13 Asymptotic Equipartition Principle

Consider an iid sequence: $X_1\ X_2\ \ldots$ iid $\sim p_X(x)$

Typical Set: $A_\varepsilon^{(n)} = \left\{ x^n : \left| \frac{1}{n}\log\frac{1}{p(x^n)} - H(X) \right| \le \varepsilon \right\}$ ← Depends on $p_X(x)$.

$\varepsilon > 0$ "typical sequence"

Sample entropy

$p(x^n) = \prod_{i=1}^{n} p_X(x_i)$ because iid.

$\frac{1}{n}\log\frac{1}{p(x^n)} = \frac{1}{n}\sum_{i=1}^{n}\log\frac{1}{p_X(x_i)} \xrightarrow{LLN} E\log\frac{1}{p_X(X)} = H(X)$

Properties:
1.) $P(A_\varepsilon^{(n)}) \longrightarrow 1$ as $n \to \infty$
2.) $|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$
3.) $|A_\varepsilon^{(n)}| \ge 2^{n(H(X)-\varepsilon')}$ for any $\varepsilon' > \varepsilon$ and n large enough.

Proofs:
1.) LLN
2.) $1 \ge P(A_\varepsilon^{(n)})$

$= \sum_{x^n \in A} p(x^n) \ge \sum_{x^n \in A} 2^{-n(H(X)+\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)+\varepsilon)}$ □

3.) For n large enough, $P(A_\varepsilon^{(n)}) \ge 2^{-n(\varepsilon'-\varepsilon)}$

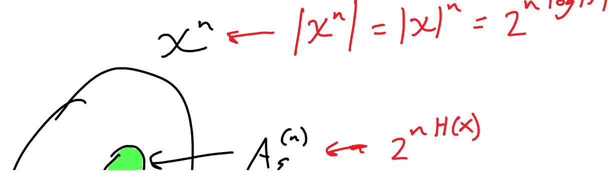
$P(A) = \sum_{x^n \in A} p(x^n) \le \sum_{x^n \in A} 2^{-n(H(X)-\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)-\varepsilon)}$ □

Notice: "Almost everything is almost equally likely."

$2^{-n(H(X)+\varepsilon)} \le p(x^n) \le 2^{-n(H(X)-\varepsilon)} \quad \forall x^n \in A_\varepsilon^{(n)}$

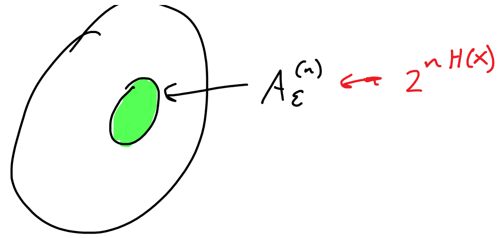
Application: Lossless source coding.

Only encode (enumerate) the typical set.



Only encode (enumerate) the typical set.
Let the rest be error.

$$\# \text{bits} = \log_2 |A_\varepsilon^{(n)}| \le n\left(H(X)+\varepsilon\right)$$



Jointly Typical:

$$A_\varepsilon^{(n)}(X,Y) = \left\{ \begin{array}{l} (x^n, y^n) : \left| \frac{1}{n} \log \frac{1}{p(x^n,y^n)} - H(X,Y) \right| \le \varepsilon \\ \text{and } x^n \in A_\varepsilon^{(n)} \text{ and } y^n \in A_\varepsilon^{(n)} \end{array} \right\}$$

iid $\sim P_{X,Y}(x,y)$

Properties:

1.) $P(A_\varepsilon^{(n)}) \to 1$ as $n \to \infty$

2.) $|A_\varepsilon^{(n)}(X,Y)| \le 2^{n(H(X,Y)+\varepsilon)}$

3.) $|A_\varepsilon^{(n)}(X,Y)| \ge 2^{n(H(X,Y)-\varepsilon')}$ for any $\varepsilon' > \varepsilon$ and n large enough.

Same proofs.

4.) $P_{\prod P_X P_Y}\left(A_\varepsilon^{(n)}(X,Y)\right) \le 2^{-n(I(X;Y)-3\varepsilon)}$

5.) $P_{\prod P_X P_Y}\left(A_\varepsilon^{(n)}(X,Y)\right) \ge 2^{-n(I(X;Y)-3\varepsilon')}$ for any $\varepsilon' > \varepsilon$ and n large enough.

Proofs: $P_{\prod P_X P_Y}\left(A_\epsilon^{(n)}(X,Y)\right) = \sum_{x^n y^n \in A} p(x^n)p(y^n)$

4.) $p(x^n) \le 2^{-n(H(X)-\varepsilon)}$ $\quad p(y^n) \le 2^{-n(H(Y)-\varepsilon)}$ $\quad \forall\, x^n \in A,\ y^n \in A$

$$\Rightarrow \sum_{x^n y^n \in A} p(x^n)p(y^n) \le \sum_{x^n y^n \in A} 2^{-n(H(X)+H(Y)-2\varepsilon)}$$

$$= |A_\varepsilon^{(n)}(X,Y)|\, 2^{-n(H(X)+H(Y)-2\varepsilon)}$$

$$\le 2^{n(H(X,Y)+\varepsilon)}\, 2^{-n(H(X)+H(Y)-2\varepsilon)} \quad \square$$

5.) Use opposite bounds.

Classical Channel Capacity: Achievability Proof:

Let $R < I(\mathcal{N}^{X \to Y})$

For any $\varepsilon > 0$, construct a communication system at rate R that achieves error prob. $\leq \varepsilon$ over the memoryless channel defined by $\mathcal{N}$.

(i.e. Design n. $f: [2^{nR}] \to \mathcal{X}^n$, and $g: \mathcal{Y}^n \to [2^{nR}]$)

blocklength — encoder — Decoder

Proof: Choose $\bar{P}_X$ such that $R < I(X;Y)$ evaluated according to $\bar{P}_X P_{Y|X}$

Channel.

- Let $\mathcal{C} = \{x^n(m)\}_{m=1}^{2^{nR}}$ represent a "codebook" for transmission.

  Encoder: $f(m) = x^n(m)$

- Use a "joint typicality decoder":

  Suppose y^n is observed:
  - If there is a unique m' such that $(x^n(m'), y^n) \in A_{\varepsilon'}^{(n)}$, Output $\hat{M} = m'$
  - If there is more than one jointly typical codeword or none, then error (output $\hat{M} = 1$).

Now we must select the blocklength n, the ε' for the typical set (decoder), and the codebook.

- n will be "large enough"
- ε' will depend on the gap $I(X;Y) - R$.
- Existance of a good $\mathcal{C}$ will be shown by Shannon's random proof technique.

Random Codebook:

Generate $x^n(m)$ iid for each m $\sim \prod_{i=1}^{n} \bar{P}_X(x_i)$

Consider expected performance.

Step 1:

Consider average error over messages.

Def: $\lambda_m = \Pr(\mathcal{E} \mid M = m)$ ← Depends on encoder and decoder $\Rightarrow$ depends on $\mathcal{C}$.

2^{nR}

Show $\frac{1}{2^{nR}} \sum_{m=1} \lambda_m$ is small.

$$E_{\mathcal{C}} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \lambda_m = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} E_{\mathcal{C}} \lambda_m$$

$$= \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} E_{\mathcal{C}} \lambda_1 \qquad \text{by symmetry}$$

$$= E_{\mathcal{C}} \lambda_1$$

$$= E_{\mathcal{C}} \Pr(\mathcal{E} \mid M=1)$$

$$\leq E_{\mathcal{C}} \Pr\Big((X^n(1), Y^n) \notin A_\varepsilon^{(n)} \text{ or } \exists m' \neq 1 \text{ s.t. } (X^n(m'), Y^n) \in A_\varepsilon^{(n)} \;\Big|\; M=1 \Big)$$

$$\leq E_{\mathcal{C}} \Pr\big((X^n(1), Y^n) \notin A_\varepsilon^{(n)} \mid M=1 \big) \longleftarrow \text{Call this } \alpha_1$$

$$+ \sum_{m'=2}^{2^{nR}} E_{\mathcal{C}} \Pr\big((X^n(m'), Y^n) \in A_\varepsilon^{(n)} \mid M=1 \big) \longleftarrow \alpha_2$$

$\alpha_1 \to 0$ by Prop. 1 of J.T.

$\alpha_2 \leq 2^{nR} P_{\prod P_X P_Y} \left(A_{\varepsilon'}^{(n)} \right) \leq 2^{-n(I(X;Y) - R - 3\varepsilon')}$

$\Rightarrow$ Choose $\varepsilon' < \frac{1}{3}(I(X;Y) - R)$ and let n be large enough.

Notice $E_{\mathcal{C}} \frac{1}{2^{nR}} \sum_m \lambda_m$ becomes arbitrarily small with n.

<u>Step 2:</u> Exists a codebook that performs as well as the expected value.

<u>Step 3:</u> Average error $\to$ Maximum error.

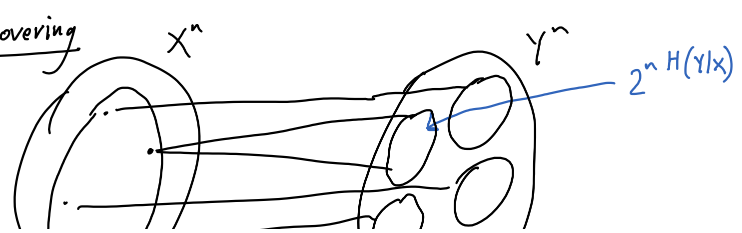
Notice $\lambda_m \geq 0 \Rightarrow \lambda_m \leq 2 \frac{1}{2^{nR}} \sum_{m'=1}^{2^{nR}} \lambda_{m'}$ for at least half of the $m \in [2^{nR}]$

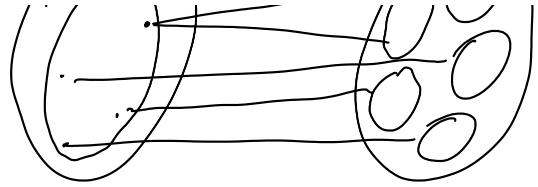
Markov inequality.

Only use those messages. Throw out the worst half of the message.

Messages $= \frac{1}{2} 2^{nR} = 2^{n(R - \frac{1}{n})}$ which has a negligible effect on rate.

<u>Under-covering</u>





$$2^{nH(Y|X)} 2^{nR} \leq 2^{nH(Y)}$$

$$R \leq I(X;Y)$$

Strong Typical Set:

Define empirical distribution:

Let $N(k|x^n) = \sum_{i=1}^{n} \mathbb{1}(x_i = k)$ ← Count the number of occurances of k

Example: $N(1|011000) = 2$

Empirical distribution ("type"):

$$t_{x^n}(x) = \frac{1}{n} N(x|x^n)$$

Notice that $t_{x^n}(x)$ is a pmf. (i.e. $t_{x^n}(x) \geq 0$ and $\sum_x t_{x^n}(x) = 1$)

Example: $t_{011000}(x) = \begin{cases} 2/3, & x=0 \\ 1/3, & x=1 \end{cases}$

Strongly typical set:

Defines the typical set.

$$T_\varepsilon^{(n)} = \{ x^n : \forall x \in \mathcal{X}, \ |t_{x^n}(x) - p_X(x)| \leq \varepsilon p(x) \}$$

Definition in text:

$$T_\varepsilon^{(n)} = \left\{ x^n : \forall x \in \mathcal{X}, \ |t_{x^n}(x) - p_X(x)| \leq \varepsilon \text{ and } t_{x^n}(x) = 0 \text{ if } p(x) = 0 \right\}$$

Properties:

1.) $P(T_\varepsilon^{(n)}) \longrightarrow 1$ as $n \to \infty$ LLN.

2.) $\forall x^n \in T_\varepsilon^{(n)}$, $\forall g$: $(1-\varepsilon) E_p g(X) \leq E_{t_{x^n}} g(X) \leq (1+\varepsilon) E_p g(X)$

2.) $\forall x^n \in T_\varepsilon^{(n)}$: $(1-\varepsilon) E_p g(X) \le E_{t_{x^n}} g(X) \le (1+\varepsilon) E_p g(X)$
$\forall g$

3.) $T_\varepsilon^{(n)} \in A_{\varepsilon H(X)}^{(n)}$ ← Proven from Prop. 2.

i.e. inherets all properties of weak typicality.