STO8 LECTURE 2 2/11/08

Note Title 2/11/2008

TODAY: SHANNON'S PAPER

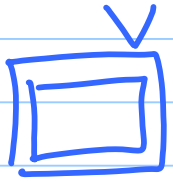
- OVERVIEW
- NOISY CODING THEOREM (for BSC)
- CONVERSE

OVERVIEW

Shannon's entities

← Source of Information

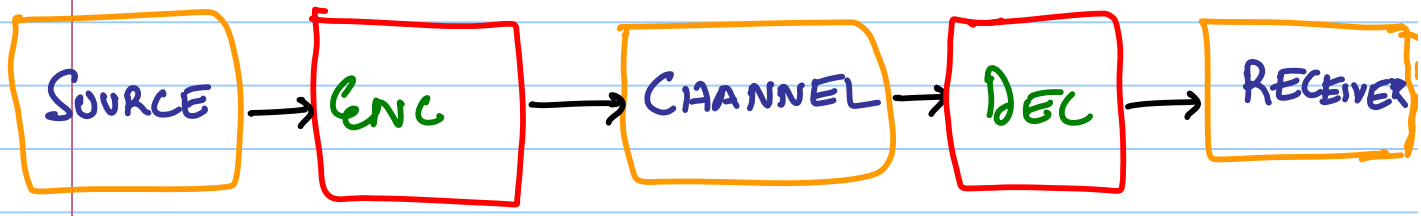
← Channel of Communication



← Receiver

Contributions

- Architecture of Information Transmission System.
  (use of ENCODER + DECODER)



- Modelling Source & Channel (Mathematically)

  Source: Stochastic Process ≡ Distribution
  Quantified by Rate at which
  it produces uncertainty
  (leads to Entropy)

Channel: Modelled by Input → Output Process ; "Marginal Distribution")

Quantified by Mutual Information

⇒ **Capacity** of Channel

Meta-Theorem:

- if Rate < Capacity then can transmit reliably

$$Pr[\text{error}] = \exp(-n)$$

- If Rate > Capacity then can not trasmit reliably

$$Pr[\text{correct transmission}] = \exp(-n)$$

where n = length of transmission.

Further Details

Noiseless Coding Theorem

Case of Noiseless Channel; Input → Output is Identity function;

Need to compress information produced by source; Can be done ...

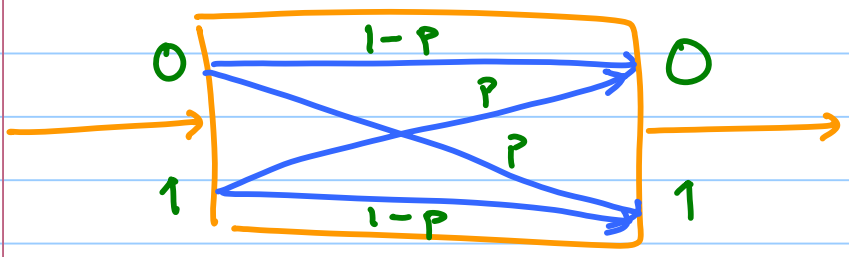
Noisy Coding Theorem

Assume source is uniform on domain (no compression possible)

Need to add redundancy... Can be done.

Meta-Theorem: Obtained by putting Noiseless + Noisy Theorems together.

Example: Binary Symmetric Channel



Shannon's Theorem: Capacity = $1 - H(p)$

where $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$

Explanation: Where does $H(p)$ come from?

Suppose we transmit $X_1, \ldots X_n$ thru channel

Receive $X_1+Y_1, \ldots, X_n+Y_n$ where

$Y_i = 1$ w.p. p

$= 0$ w.p. $1-p$

Distribution of $Y_1 \dots Y_n = ?$

Chernoff Bounds: $Y_1 \dots Y_n$ i.i.d.

$$E[Y_i] = p \; ; \quad Y_i \in [0,1]$$

$$\Rightarrow \Pr\left[\left|\frac{\sum Y_i}{n} - p\right| > \epsilon\right] \leq e^{-\epsilon^2 n}$$

"Typical" Error Pattern: Has about pn [between $(p-\epsilon)n$ & $(p+\epsilon)n$] bit flips.

Size of set $\approx \binom{n}{pn} \approx 2^{H(p)\cdot n}$

This is where $H(p)$ enters.

formally

$$\leq 2\epsilon n \cdot \binom{n}{(p+\epsilon)n} \leq 2^{(H(p)+\delta)\cdot n}$$

where $\delta \to 0$ as $\epsilon \to 0$.

Moreover: Prob. of any "typical" error is about the same (by symmetry) & equals $\approx 2^{-H(p)\cdot n}$

Formally: if $y = y_1 \dots y_n$ has $i \in [(p-\epsilon)n, (p+\epsilon)n]$ ones

$$\Pr[Y_1 \dots Y_n = y] \leq \frac{1}{\binom{n}{i}}$$

$$\leq 2^{-(H(p)-\delta)n}$$

Suffices to prove Coding theorem + converse, for BSC(p).

Coding Theorem [in gory detail]

$\forall p \in (0, \frac{1}{2})$, $\epsilon > 0$ $\exists \delta > 0$, n_0, s.t. $\forall n \geq n_0$

$\exists k$, $E : \{0,1\}^k \to \{0,1\}^n$

$D : \{0,1\}^n \to \{0,1\}^k$

s.t.

(1) $\Pr_{\substack{m \leftarrow_U \{0,1\}^k \\ y = y_1 \ldots y_n \leftarrow BSC(p)}} \left[D(E(m) + y) \neq m \right] \leq 2^{-\delta n}$

(2) $k \geq (1 - H(p+\epsilon)) \cdot n$

Proof: (Will specify δ, n_0 later ... Let $k = \lceil (1 - H(p+\epsilon)) \cdot n \rceil$)

★ Need to specify E, D, and show they work ★

Choice of E: Pick $E : \{0,1\}^k \to \{0,1\}^n$ at random.

$(\forall\, m_1, m_2$

$E(m_1)$ uniform over $\{0,1\}^n$

$E(m_2)$ ind. of $E(m_1))$

Choice of D, given E:

$D(x) =$ if $\exists$ unique $m \in \{0,1\}^k$ s.t.

$\Delta(E(m), x) \leq (p + \frac{\epsilon}{2})n$,

then output m,

else output ERROR.

Analysis :

Bad Events

(1) E_1 : Too many errors

$$\frac{\sum Y_i}{n} - p > \frac{\epsilon}{2}.$$

(2) E_2 : Bad Error pattern (or message m)

$$\exists\, m' \neq m, \quad \Delta\left(E(m'), E(m) + Y\right) \leq \left(p + \frac{\epsilon}{2}\right) n$$

Will show

A. $\Pr[E_1] \longrightarrow \exp(-n)$

B. $\Pr[E_2] \rightarrow \exp(-n)$

C. $\neg E_1$ and $\neg E_2 \Rightarrow$ Decoding correct.

(C) : Follows from definition.

(A) : Chernoff bounds $\Rightarrow$

$$\Pr[E_1] \leq e^{-\frac{\epsilon^2}{4} \cdot n}$$

(B) : Fix $E(m)$, Y, $m' \neq m$ & Pick $E(m')$ at random

$$\Pr_{E(m')}\left[\Delta(E(m'), E(m)+Y) \leq (p+\tfrac{\epsilon}{2})n\right]$$

$$\leq \frac{2^{H(p+\frac{\epsilon}{2}) \cdot n}}{2^n}$$

$$\approx 2^{-(1-H(p+\frac{\epsilon}{2}))n}$$

$$\Pr\left[\exists\, m' \neq m \text{ s.t. } \ldots\right]$$

$$\leq 2^k \cdot 2^{-(1-H(p+\frac{\epsilon}{2}))n}$$

$$= 2^{[H(p+\frac{\epsilon}{2}) - H(p+\epsilon)] \cdot n} \to 2^{-\Theta(n)}$$

Theorem follows by choosing δ, n_0, etc. carefully.

Converse: Is rate $k \approx (1 - H(p))n$ best possible? Could we have chosen better? Shannon's answer: **No!**

Converse Coding Theorem:

$\forall p, \epsilon > 0, \quad \exists \delta > 0, n_0 \quad \forall n \geq n_0,$

$\forall k, E, D \text{ s.t. } k \geq (1 - H(p-\epsilon))n$

$E: \{0,1\}^k \to \{0,1\}^n$

$D: \{0,1\}^n \to \{0,1\}^k$

$$\Pr_{m,y}\left[D(E(m)+y) = m\right] \leq 2^{-\delta n}.$$

Proof: Intuition: # Typical Errors = $2^{H(p)n}$.

Correct recovery $\Rightarrow$ Can determine message & error!

But # errors $\times$ # messages $\gg 2^n$

Formal Proof: Notation: m = transmitted word
y = error
$wt(y)$ = # errors.
$= \sum y_i$

Bad Events

E1: # errors too small

$$wt(y) \leq \left(p - \frac{\epsilon}{2}\right) n$$

E2: error is too nice ....

$\exists\ x$ s.t.

$$wt(x - E(D(x))) \geq \left(p - \frac{\epsilon}{2}\right) n$$

& E2(x): $m = D(x)$
$y = x - E(D(x))$

(A): Neither E1 nor E2 occur ⇒ Decoding Error.

Let $x = E(m) + Y$

Decoding Correct ⇒ $D(x) = m$

E1: ⇒ $wt(Y) = wt(x - E(D(x))$ ✗

$\geq (\rho - \frac{\epsilon}{2})n$

E2 ⇒ $E2(x) \Rightarrow$

$m \neq D(x)$

Or $Y \neq x - E(D(x))$

Either way contradiction

(B) Chernoff Bounds

$$\Rightarrow \Pr_{y}\left[wt(y) \leq \left(p-\frac{\epsilon}{2}\right)n \right] \leq e^{-\frac{\epsilon^2}{4}\cdot n}.$$

(C) $E2(x)$ given $\Delta(x, E(D(x)) \geq \left(p-\frac{\epsilon}{2}\right)n$

$$\Pr_{m}\left[m = D(x) \right] = 2^{-k}$$

$$\Pr_{y}\left[y = x + E(D(x)) \right] \leq \frac{1}{\binom{n}{\left(p-\frac{\epsilon}{2}\right)n}}$$

$$= 2^{(-H(p)+\delta)n}$$

$$\Pr\left[E2(x) \right] \leq \approx 2^{-k} \cdot 2^{-H(p)\cdot n}$$

$$\Pr[E2]$$

$$= \Pr\left[\exists x \text{ s.t. } wt(x - E(D(x))) \geq \left(\rho - \frac{\epsilon}{2}\right)n \;\&\; E2(x)\right]$$

$$\leq \sum_x \Pr[E2(x)]$$

$$\leq\approx 2^{n} \cdot 2^{-k} \cdot 2^{-\mathcal{H}(\rho)\cdot n} \approx 2^{-\delta n} \ldots$$

Conclusions

Shannon Theory: Dramatically powerful.
Captures many interesting cases,
much richer than BSC.

But doesn't imply perfect understanding:

Example: The Deletion Channel (p).

Transmit n bits (asynchronously)
Each bit received w.p. $1-p$
& not received at all w.p. p

Shannon's Theorem $\exists$ $C = C(p)$ s.t.
Can transmit at rate $< C$ but not higher. But what is $C(p)$?

--- Not known ...

(Question raised in e.g.

[Driencu, Mitzenmacher IT 2007].)