ST06 - LECTURE 02

2/9/6

Note Title 2/8/2006

Today

- Entropy
- Mutual Information

Entropy "Measures average randomness in a random variable"

Example: $X \in \{0,1\}$

$$\Pr[X=0] = 1/2 = \Pr[X=1]$$

$$Y \in \{0,1\}$$

$$\Pr[Y=0] = 7/8$$

$$\Pr[Y=1] = 1/8$$

Which one is more random?

$$Z \in \{0, 1, 2\}$$

$$\Pr[Z=0] = 9/10$$

$$\Pr[Z=1] = 1/20$$

$$\Pr[Z=2] = 1/20$$

Now which is more random?

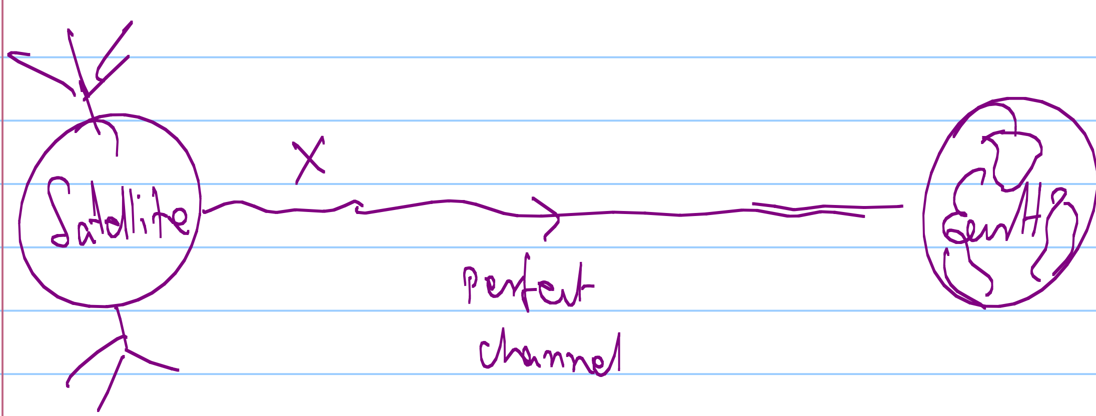
Intuition: X is more random than Y

But Y vs. Z .... less clear!

Problem: - Don't have formal reason why we think X is more random than Y

- if we did, we might be able to say something about Y vs. Z.

Lets think harder about X vs. Y. & lets put on our engineer's hat.



X takes one bit to comm.

Y takes one bit to comm.

Conclusion: X as random as Y ?

Last lecture: This is not the right way to think about it !!

Should collect n symbols

$X_1 \dots X_n$ i.i.d dist as X,

$Y_1 \dots Y_n$ i.i.d dist as Y

& compare expected length of transmission.

So can we compute this length ?
will length$/n$ have nice behavior ? Lets see.

A Generic Encoding Scheme for

$\mathrm{Enc}(Z_1 \ldots Z_n)$: $Z_i \in \{0,1\}$ i.i.d.

$\Pr[Z_i = 1] = p$

First send $k = \sum Z_i$.

then send $\left\lceil \log_2 \binom{n}{k} \right\rceil$ bits to explain which of the $\binom{n}{k}$ possibilities occurred.

let $\|\mathrm{Enc}(Z_1 \ldots Z_n)\|$ denote its encoding length;

$$\mathop{E}_{Z_1 \ldots Z_n}\left[\|\mathrm{Enc}(Z_1 \ldots Z_n)\|\right]$$

$$= \sum_{k=0}^{n} \Pr\left[\sum z_i = k \right] \cdot \left\lceil \log_2 \binom{n}{k} \right\rceil$$

"Chernoff-bounds" says

$$\Pr\left[\sum z_i \notin \left[(p-\epsilon)n, (p+\epsilon)n \right] \right] \le 2^{-\epsilon^2 n}$$

$$\Rightarrow E\left[\| \mathrm{Enc}(z_1 \dots z_n) \| \right]$$

$$= \sum_{k=(p-\epsilon)n}^{(p+\epsilon)n} \Pr\left[\sum z_i = k \right] \cdot \left\lceil \log_2 \binom{n}{k} \right\rceil \pm 2^{-\epsilon^2 n} \cdot n$$

- But now $\log_2 \binom{n}{k} = \log_2 \binom{n}{pn} \pm \delta n$

$$(\delta \to 0 \text{ as } \epsilon \to 0)$$

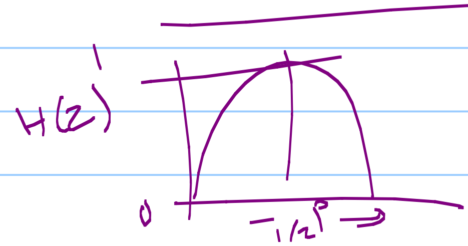
- $$\binom{n}{pn} \approx \frac{n^n}{(pn)^{pn} \cdot ((1-p)n)^{(1-p)n}} \approx \left(\frac{1}{p}\right)^{pn} \cdot \left(\frac{1}{(1-p)}\right)^{(1-p)n}$$

$$\log_2 \binom{n}{pn} \approx n \underbrace{\left[p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \right]}_{H(z)}$$

Conclude:

Associate charge of $-\log_2 p$
for transmitting 1
& charge of $-\log_2 (1-p)$
for transmitting 0

& total charge $\approx$ expected length of transmission.



Convex function.

After all this work ... figured out the "randomness" / "Entropy" of a binary random variable.

What about non-binary values?

X taking values in $\Omega = \{1 \dots N\}$

let $p_i = \Pr[X = i]$

Consider $Y = 1$ if $X = 1$

$= 0$ o.w.

$$H(Y) = -\Big[p_1 \log p_1 + (1-p_1) \log (1-p_1) \Big]$$

Now consider expected encoding length (averaged over n copies) of X :

First send Y

then send $X \mid Y$

"Cost of Y" $= H(Y)$

Cost of $X \mid Y = ?$

if $Y=1 \Rightarrow$ cost $=0$

if $Y=0 \Rightarrow$ cost $= H(\tilde{X})$

where $\tilde{X} \in \{2, \ldots, N\}$

$$\Pr[\tilde{X} = i] = \frac{p_i}{1-p_1}$$

Expected cost of encoding X

$$H(X) = H(Y) + (1-p_1) H(\tilde{X})$$

$$= -\Bigg[p_1 \log p_1 + (1-p_1)\log(1-p_1)$$

$$\sum \frac{p_i}{} \log \frac{p_i}{(1-p_1)} \Bigg]$$

$$= -\Bigg[p_1 \log p_1 + (1-p_1)\log(1-p_1)$$

$$+ \frac{1}{} \sum_{i=2}^{N} p_i \log p_i$$

$$- \sum_{i=2}^{N} p_i \log(1-p_1) \Bigg]$$

$$= -\left[\sum p_i \log p_i \right]$$

Conclusion : Defn of $H(X)$

X takes on values $x_1 \dots x_N$

w.p. $p_1 \dots p_N$

then
$$H(X) = H(p_1, \dots, p_N)$$
$$= -\sum_{i=1}^{N} p_i \log_2 p_i .$$

Examples :

- $Y = 1$ w.p. $1/8$

  $= 0$ w.p. $7/8$

$$H(Y) = \frac{1}{8} \log_2 8 + \frac{7}{8} \log_2 \frac{8}{7}$$
$$\approx .5436$$

- $Y = H$ w.p. $\frac{1}{2}$

  $= TH$ w.p. $\frac{1}{4}$

  $= TTH$ w.p. $\frac{1}{8}$

  $\vdots$

- $H(Y) = ?$

$$= \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot k$$

$$= 1$$

- $Y = AA$ w.p. $(1-p)^2$

  $= AB$ w.p. $(1-p)\,p$

  $= BA$ w.p $(1-p)\,p$

  $= BB$ w.p. p^2

$H(Y) = 2H(p)$. (Suprised ?)

Essential properties of $H(p_1, \ldots, p_N)$

1. H is a symmetric function of its arguments

2. $H(p_1, \ldots, p_N) \leq \log_2 N$

3. $$H(p_1, \ldots, p_N) = H(p_1, 1-p_1) + (1-p_1)\, H\left(\frac{p_2}{1-p_1}, \ldots, \frac{p_N}{1-p_N}\right)$$

Turns out our defn. is the <u>only</u> one that satisfies 1, 2, 3.

- Joint Entropy

(X, Y) distributed jointly : Can define entropy of the joint variable

$$H(X,Y) = \sum_{\substack{x \in \Omega_X \\ y \in \Omega_Y}} P(x,y) \log_2 \frac{1}{P(x,y)}$$

- Conditional Entropy

Suppose (X, Y) distributed jointly & not independent.

E.g. $X \in \{0, 1\}$ w.p. $\frac{1}{2}$

$$\begin{aligned} Y &= X && \text{w.p. } 1-p \\ &= \bar{X} && \text{w.p. } p \end{aligned}$$

What is the entropy in X given Y

(e.g. X transmitted, Y received)

How much uncertainty about x?

Let X_y denote the r.v. $X \mid \{Y = y\}$

then $$H(X \mid Y) = \sum_{y \in \Omega_Y} P_Y(y) \cdot H(X_y)$$

marginal dist on Y.

in above example

$$H(X \mid Y) = H(p)$$

Chain Rule of Entropy

$$H(X, Y) = H(X) + H(Y \mid X)$$

Proof: Calculation

Next fact

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

$$\Rightarrow \underbrace{H(X) - H(X|Y)}_{} = H(Y) - H(Y|X)$$

Let's call this $I(X,Y)$: Mutual Information

Question: Is mutual information positive?

i.e. $H(X) \geq H(X|Y)$?

Intuitively : YES

But can't push our luck like this indefinitely. Should actually prove this.

$\boxed{\text{NEXT LECTURE}}$