Discrete Info Sources & Entropy

Info content of a message is related to amount of "surprise" conveyed by message

Info content $\equiv$ surprise conveyed by message

Ex: Capital of Austria is Vienna

Capital of Nigeria is ?

First time you heard this ~~info~~ it was <u>informative</u>

~~2nd time~~ If you hear it again, it is data, but holds no information (like reloading a page)

Info is distinguished by property that it <u>adds</u> to our knowledge

<u>Ex:</u> Tx sending

00000000...

It is sending data, but no information

<u>Source:</u>

defined by:

- set of <u>output symbols</u> it is capable of producing : $\boxed{\text{Source Alphabet}}$
- probability rule that <u>govern</u> the emission of the symbols : $\boxed{\text{Probability Rule}}$

Source alphabet with M possible symbols

$$A = \{a_0, a_1, \dots, a_{M-1}\}$$

- Source outputs symbol in a time sequence

$$\bar{a} = (s_0, s_1, \dots, s_t \dots)$$

$s_i \in A$ symbol emitted at time t.

- At a given time index, prob. that source emits symbol a_m is

$$p_m = P(a_m)$$

- ~~If the set of~~ This creates set of probabilities

$$P_A = \{p_0, p_1, \dots, p_{M-1}\}$$

- If the set of probabilities is not a func. of time, the source is said to be stationary

- By the axiom of prob.

$$\sum p_m = 1$$

- Memoryless:

Questions:

How do we deal with the following two cases

- The alphabet that the source emits are different in each instant?

- We deal with synchronous sources? What if the source is asynchronous? (time interval between emitted symbols is not fixed)

Define a null character a_M

If the source does not emit anything at time t, we say a_M is

How do we transmit these symbols to a Comm. channel?

Use binary representation

Ex: $M = 4$

$$\left.\begin{array}{l} a_0 \longrightarrow 00 \\ a_1 \longrightarrow 01 \\ a_2 \longrightarrow 10 \\ a_3 \longrightarrow 11 \end{array}\right\} \text{Source data}$$

Data ▷ Information:

Data is not equivalent to information

- Ex. source with an alphabet of one symbol only

symbol is data but not info

data is completely uninformative

info = surprise

amount of info = amount of surprise

no surprise $\Rightarrow$ no info.

How can we measure the info content of Source?

Info content is an important atribute

Can be measured

Shanon in one of his original papers gave a math def. for the

Entropy: avg amount of info conveyed per source symbol of source

$$H(A) = \sum_{m=0}^{M-1} p_m \log(1/p_m) \quad \text{(bit)}$$

log to base 2 (not if we use natural log)

Entropy defined by individual symb. probabilities

$$= E \log(1/p_i)$$

For uninformative source

$M = 1 \Rightarrow p_0 = 1$

$\Rightarrow H(A) = 0.$

Ex

What is the entropy of a 4 - Source $\{a_0, a_1, a_2, a_3\}$ having the symbol probabilities

$$P_A = \{0.5, 0.3, 0.15, 0.05\}$$

$$H(A) = 0.5 \log_2(2) + 0.3 \log_2(10/3) + 0.15 \log_2(100/15) + 0.05 \log_2(100/5)$$

$= 1.6477$ bits (avg. amount of info per symb.)

We represent each symbol as follows

a_0	00
a_1	01
a_2	10
a_3	10

What is the avg ~~bit~~ tx per symbol?

$2 \times 0.5 + 2 \times 0.15 + 2 \times 0.3 + 2 \times 0.05$

$= 2 \times 1$

$= 2$ bits (avg. no of bits Tx per symbol)

What is the info efficiency ?

$$\text{Info Efficiency} = \frac{1.6477}{2}$$
$$\approx 82.4\%$$

Approximately 17.6 % of bits transmitted are wasted

How can we increase the efficiency?

1. What if the source symbols are equiprobable?

$$P_A' = \{1/4, 1/4, 1/4, 1/4\}$$

In this case

$$H(A') = (1/4 \log_2 4) \times 4$$
$$= 2 \text{ bits}$$

$$\Rightarrow \text{Info efficiency} = \frac{2}{2} = 100\%$$

2. Have a different bit assignment (source encoder)

$a_0 \longrightarrow 0$

$a_1 \longrightarrow 11$

$a_2 \longrightarrow 101$

$a_3 \longrightarrow 111$

$H(A) = 1.6477$ (still the same)

$$\bar{L} = \text{Avg. no of bits Tx per ~~symbol~~ codeword} = 1 \times 0.5 + 2 \times 0.15 + 3 \times 0.3 + 3 \times 0.05$$
$$= 1.7$$

$$\text{Info Efficiency} = \frac{1.6477}{1.7}$$
$$= 96.9\%$$ (much better)

$\Rightarrow$ only 3% of bits are wasted.

Can we reduce $\bar{L}$ ever further?

Use the assignment

$a_0 \longrightarrow 0$

$a_1 \longrightarrow 1$

$a_2 \longrightarrow 01$

$a_3 \longrightarrow 10$

$\bar{L} = ?$

Is there a problem with the last 2 encoders?

<u>Generalization of pt. 1:</u>

Given an M-ary source:

What dist. of prob. P_A maximizes the info content of A.

Need to find p_m to maximize

$$H(A) = \sum_{m=0}^{M-1} p_m \log_2(1/p_m)$$

~~we also have~~ subject to the constraint

$$\sum_{m=0}^{M-1} p_m = 1$$

To solve this, we write the entropy in the equivalent form

$$f(p_0, \ldots, p_{M-1}) = \frac{1}{\ln(2)} \sum_{m=0}^{M-1} p_m \ln(1/p_m) + \lambda \left(\sum_{m=0}^{M-1} p_m - 1 \right)$$

λ is what is called Lagrange multiplier

Maximize by diff. f with respect to p_i & setting result to zero to get

$$\frac{\partial f}{\partial p_i} = \frac{1}{\ln(2)} \left[\ln\left(\frac{1}{p_i}\right) - 1 \right] + \lambda$$

$$\Rightarrow \ln\left(\frac{1}{p_i}\right) = 1 - \lambda \ln(2) \quad \text{(indep. of } i\text{)}$$

$$\Rightarrow p_1 = p_2 = \cdots = p_j = 1/M.$$

Does result make sense?

- Each symbol in A is equally probable

$\Rightarrow$ observe has no idea what symbol will be emitted next

$\Rightarrow$ Each symbol has max. surprise value

$\Rightarrow$ avg. amount of info is similarly maximized

What happens if one symbol has a larger prob?

I can bet on that symbol & make money

$\Rightarrow$ no surprise

So, for an ~~s~~ M-ary source, the entropy satisfies

$$\boxed{0 \leq H(A) \leq \log_2 M}$$

Can you build an encoder that is decodable?

Excercise

$a_0 \rightarrow 0$

$a_1 \rightarrow 10$

$a_2 \rightarrow 110$

$a_3 \rightarrow 111$

No code word is a prefix of a longer code word $\Rightarrow$ instantaneously decodable

A prefix code

It is self punctuating

How do you decode? 110101110101000

The English language is not self punctuating

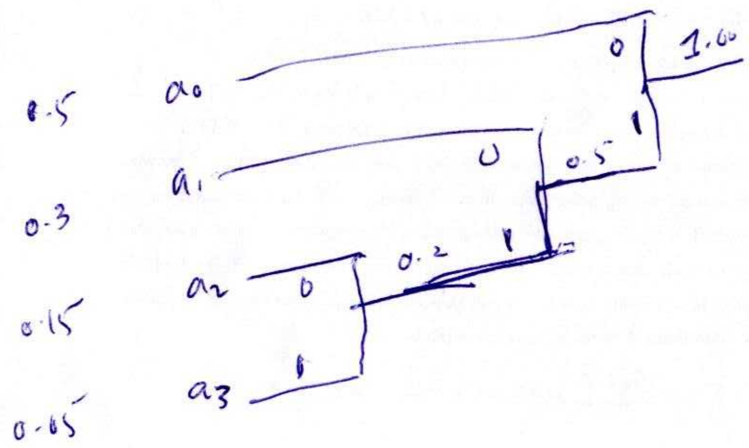
If I wanted to pick one

If I want ED to pick one

If code is not self punctuating, we need to punctuation marks.

How do we do this formally?

Huffman coding



Code new
0
10
110
111

Source Encoding

Source

Symbol out of the source

Each symbol will be either Tx or stored

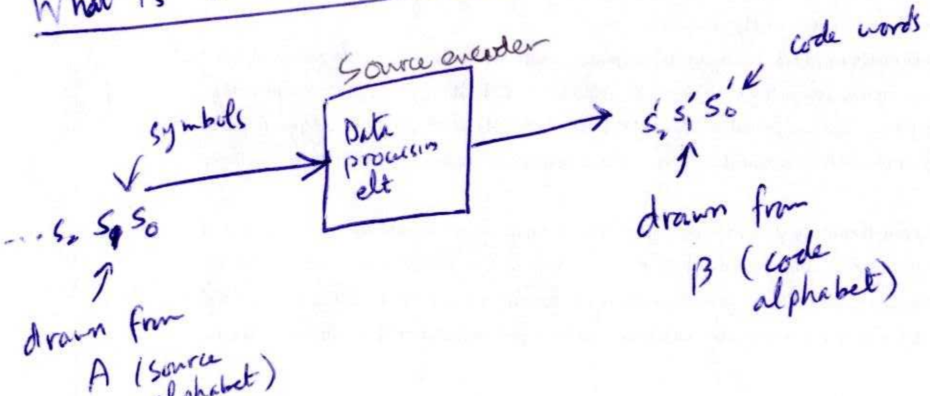
Cost associated with that

If source is inefficient

$H(A) < \log_2 |A|$

we need to make source more efficient through source coding

What is a source encoder?



Obj of source encoder

to process input such that avg. info tx (or stored) per channel use approaches $H(A)$

$C: A \rightarrow B$

Since encoded symbol must be decoded, ~~so stor~~

C must be invertible.

$\boxed{C \text{ should be one-to-one}}$

How does processing affect Information:

How different is the info. carried by the source symbols & that carried by the code words to

a

$P_0 = 0.5$	$C(a_0) \rightarrow 0$
$P_1 = 0.3$	$C(a_1) \rightarrow 10$
$P_2 = 0.15$	$C(a_2) \rightarrow 110$
$P_3 = 0.05$	$C(a_3) \rightarrow 111$

What can we do to improve the efficiency even further?

- Create 16-ary symbols instead

$a_0 a_0$
$a_0 a_1$
$a_0 a_2$
$a_0 a_3$
$\vdots$

~~Find the entropy~~
- What is the new alphabet set?

$A \times A$

- What are the symbol probabilities?

$\{ p_0 p_0, p_0 p_1, \dots, p_4 p_4 \}$

Reason: source is memoryless

What is the entropy when we take two symbols at a time

What is the intuitive answer??

The new symbol is

$c_{ij} = (a_i, a_j)$

$$H(C) = \sum_{i=1}^{M_A} \sum_{j=1}^{M_A} p_{ij} \log(1/p_{ij})$$

$p_{ij} = p_i p_j$. why?

$$H(C) = \sum_{i=1}^{M_A} \sum_{j=1}^{M_A} p_i p_j \log(1/p_i p_j)$$

$$= \sum \sum p_i p_j \log(1/p_i) + \sum p_i p_j \log(1/p_j)$$

$$= \sum p_j \sum p_i \log(1/p_i)$$

$$= \sum p_j H(A) + \sum p_i H(A)$$

$$= 2H(A)$$