Communication Theory II

Lecture 6: Information Theory

What is information?

- Can we measure information?
- Consider the two following sentences:

1. There is a traffic jam on I5
2. There is a traffic jam on I5 near Exit 234

Sentence 2 seems to have more information than that of sentence 1.From the semantic viewpoint, sentence 2 provides more useful information.

What is information?

- It is hard to measure the “semantic” information!
- Consider the following two sentences

1. There is a traffic jam on I5 near Exit 160
2. There is a traffic jam on I5 near Exit 234

It’s not clear whether sentence 1 or 2 would have more information!

What is information?

- Let’s attempt at a different definition of information.
  - How about counting the number of letters in the two sentences:

1. There is a traffic jam on I5 (22 letters)
2. There is a traffic jam on I5 near Exit 234 (33 letters)

Definitely something we can measure and compare!

Probability and information

Australia vs Sri Lanka

AUS 239/8

END OF OVER 19 3 runs • 1 wicket	**SL: 87/6** CRR: 4.57 • RRR: 4.93

END OF OVER 26 4 runs • 1 wicket		**SL: 111/8** CRR: 4.26 • RRR: 5.37	
Lasith Malinga	4 (4b 1x4)	Xavier Doherty	6-1-19-4
Angelo Mathews	11 (20b)	Steven Smith	3-0-8-0

Lasith Malinga

FORMAT	Mat	Inns	NO	Runs	Ave
ODI	226	119	36	**567**	6.83

Muthiah Muralidaran

FORMAT	Mat	Inns	NO	Runs	Ave
ODI	350	162	63	**674**	6.80

SL win probability 5%

AUS		239/8
SL	(44.2/50 ov, T:240)	**243/9**

Sri Lanka won by 1 wicket (with 34 balls remaining)

SL Malinga 56 (47) — 132 (109) — AD Mathews 66 (62)

Surprise

Probability and information

Bermuda vs Sri Lanka, 4th Match, Group E

SL	**321/6**

END OF OVER 23 9 runs — **BMUDA: 73/8** CRR: 3.17 • RRR: 9.22

Kevin Hurdle	6 (3b 1x6)	Muthiah Muralidaran	6-0-28-1
Lionel Cann	24 (30b 2x4 1x6)	Farveez Maharoof	6-1-18-3

SL win probability 95%

SL	**321/6**
BMUDA	(24.4/50 ov, T:322) **78**

Sri Lanka won by 243 runs

No Surprise

Question

Probability of an event and information, that we can get if this event is happened?

How they related? Inverse or proportional ?

What is information?

- Information Equation

$$I(p) = -log_b(\text{p})$$

- p = probability of the event happening
- b = base

(base 2 is mostly used in information theory)

unit of information is determined by base

base 2 = bits

base 3 = trits

base 10 = Hartleys

base e = nat

American mathematician, electrical engineer, computer scientist and cryptographer known as the "father of information theory"

Shannon’s Entropy

Shannon's entropy, also known as Shannon entropy or information entropy, was introduced by Claude Shannon in his seminal work on information theory titled "A Mathematical Theory of Communication," published in 1948. Shannon's paper laid the foundation for the field of information theory and revolutionized our understanding of communication and data compression.

In his paper, Shannon addressed the fundamental questions of how information can be quantified and how efficiently it can be transmitted over a communication channel. He introduced the concept of entropy as a measure of the average amount of information contained in a random variable or information source.

Shannon's work on entropy laid the groundwork for various applications in data compression, error correction coding, cryptography, and more. It provided a formal framework for measuring and manipulating information, leading to significant advancements in communication systems and data processing technologies.

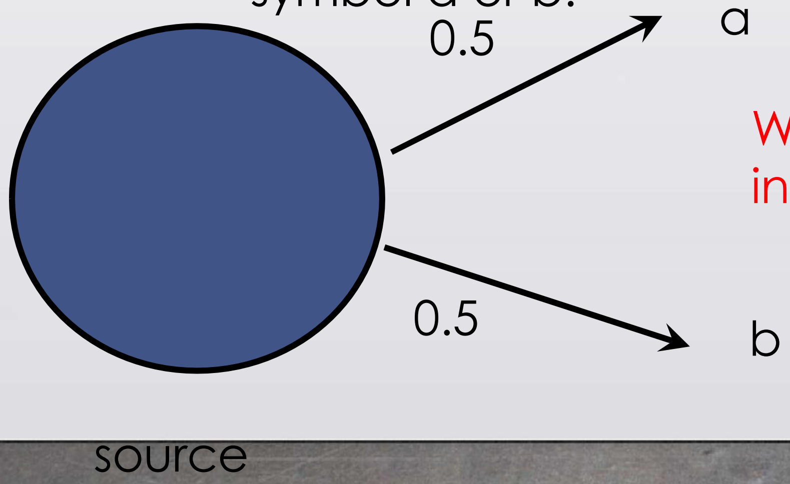
[Claude Shannon - Father of the Information Age - YouTube]

Shannon's Entropy

- Consider the following string consisting of symbols a and b:

  abaabaababbbaabbabab… ….

  - On average, there are equal number of a and b.
  - The string can be considered as an output of a below source with equal probability of outputting symbol a or b:



We want to characterize the average information generated by the source!

Self Information

- So, let's look at it the way Shannon did.
- Assume a memoryless source with
  - alphabet $A = (a_1, \ldots, a_n)$
  - symbol probabilities $(p_1, \ldots, p_n)$.
- How much information do we get when finding out that the next symbol is a_i?
- According to Shannon the self information of a_i is

$$I(a_i) = \log \frac{1}{p_i}$$

Why?

Assume two independent events A and B, with probabilities $P(A) = p_A$ and $P(B) = p_B$.

For both the events to happen, the probability is $p_A . p_B$. However, the amount of information should be added, not multiplied.

$$I(p_A \cdot p_B) = I(p_A) + I(p_B)$$

Logarithms satisfy this! $I(A) = \log(p_A)?$

No, we want the information to increase with decreasing probabilities, so let's use the negative logarithm.

$$I(A) = -\log(p_A) = \log \frac{1}{p_A}$$

Self Information

Example 1:

$$p_i = 1 \Rightarrow \quad I(1) = \log \frac{1}{1} = 0$$

Example 2:

$$p_i = 0.5 \Rightarrow I(0.5) = \log_2 \frac{1}{0.5} = 1 \text{ [bit]}$$

Which logarithm? Pick the one you like! If you pick the natural log, you'll measure in *nats*, if you pick the 10-log, you'll get *Hartleys*, if you pick the 2-log (like everyone else), you'll get *bits*.

Intuition on Shannon's Entropy

- The source output is model as discrete random variable *S*, *which takes on symbols from a fixed finite alphabet £,*

$$£ = \{s_0, s_1, \dots, s_{k\,-\,1}\}$$

With probabilities

$$P(S = s_k) = p_k \; where \;\; k = \{0,1,2, \dots, K-1\}$$

This set of probabilities must satisfy the condition

$$\sum_{k=0}^{K-1} p_k = 1$$

$$I(s_k) = 0 \qquad \text{for } p_k = 1$$

$$I(s_k) \geq 0 \qquad \text{for } 0 \leq p_k \leq 1$$

$$I(s_k) > I(s_i) \qquad \text{for } p_k < p_i$$

Self Information

On *average over all the symbols*, we get:

$$H = \sum_{1}^{N} p_i log \frac{1}{p_i}$$

$H(X)$ is called the first order *entropy* of the source.

This can be regarded as the degree of *uncertainty* about the following symbol.

More Intuition on Entropy

- Assume a binary memoryless source, e.g., a flip of a coin. How much information do we receive when we are told that the outcome is *heads*?
  - If it's a fair coin, i.e., $P(heads) = P\,(tails) = 0.5$, we say that the *amount of information is 1 bit*.
  - If we already know that it will be (or was) heads, i.e., $P(heads) = 1$, the *amount of information is zero!*
  - If the coin is not fair, e.g., $P(heads) = 0.9$, the *amount of information is more than zero but less than one bit!*
  - Intuitively, the amount of information received is the same if $P(heads) = 0.9$ or $P\,(heads) = 0.1$.

Entropy

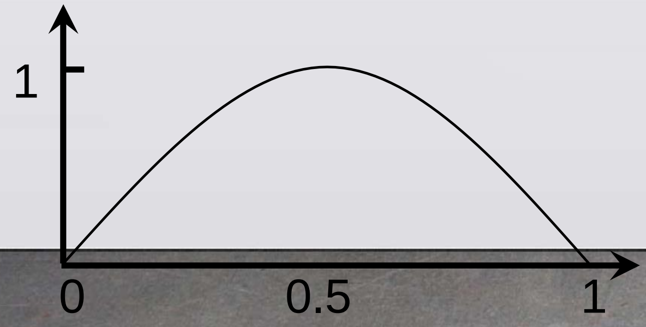
Example: Binary Memoryless Source

BMS → 0 1 1 0 1 0 0 0 …

Let $$p = P(X_k = 1)$$
$$q = P(X_k = 0) = 1 - p$$

Then $$H = p\,log\frac{1}{p} + (1-p)\,log\frac{1}{1-p}$$

Often denoted $h(p)$



The uncertainty (information) is greatest when

$$p = q = \frac{1}{2}$$

The entropy $H(\mathcal{S})$ attains its maximum value, $H_{max} = 1$ bit, when $p_1 = p_0 = 1/2$, that is, symbols 1 and 0 are equally probable.

Example

Three symbols a, b, c with corresponding probabilities:

$P = \{0.5, 0.25, 0.25\}$

What is H(P)?

Three weather conditions in Corvallis: Rain, sunny, cloudy with corresponding probabilities:

$Q = \{0.48, 0.32, 0.20\}$

What is H(Q)?

Entropy: Three properties

1. It can be shown that $0 \leq H \leq \log N$

1. Maximum entropy ($H = \log N$) is reached when all symbols are equiprobable, i.e., $p_i = 1/N$.

2. The difference $\log N - H$ is called the *redundancy* of the source.

https://www.youtube.com/watch?v=FlaJPxP8sd8&ab_channel=IainExplainsSignals%2CSystems%2CandDigitalComms

Example

Consider a discrete memoryless source with source alphabet $\mathcal{S} = \{s_0, s_1, s_2\}$ with respective probabilities

$$p_0 = \tfrac{1}{4}$$
$$p_1 = \tfrac{1}{4}$$
$$p_2 = \tfrac{1}{2}$$

yields the entropy of the source

$$\begin{aligned} H(\mathcal{S}) &= p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) \\ &= \frac{1}{4}\log_2(4) + \frac{1}{4}\log_2(4) + \frac{1}{2}\log_2(2) \\ &= \frac{3}{2} \text{ bits} \end{aligned}$$

Example

A source emits one of four possible symbols during each signaling interval. The symbols occur with the probabilities:

$$p_0 = 0.4$$
$$p_1 = 0.3$$
$$p_2 = 0.2$$
$$p_3 = 0.1$$

Find the amount of information gained by observing the source emitting each of these symbols.

s_k	s_0	s_1	s_2	s_3
p_k	0.4	0.3	0.2	0.1
$I(s_k)$ bits	1.322	1.737	2.322	3.322

Calculate the entropy of the source.

Example

Consider next the second-order extension of the source. With the source alphabet $\mathcal{S}$ consisting of three symbols, it follows that the source alphabet $\mathcal{S}^2$ of the extended source has nine symbols. The first row of Table 9.1 presents the nine symbols of $\mathcal{S}^2$, denoted as σ_0,

Alphabet particulars of second-order extension of a discrete memoryless source

Symbols of $\mathcal{S}^2$	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
Corresponding sequences of symbols of $\mathcal{S}$	s_0s_0	s_0s_1	s_0s_2	s_1s_0	s_1s_1	s_1s_2	s_2s_0	s_2s_1	s_2s_2
Probability $p(\sigma_i)$, $i = 0, 1, \ldots, 8$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

$$
\begin{aligned}
H(\mathcal{S}^2) &= \sum_{i=0}^{8} p(\sigma_i)\log_2 \frac{1}{p(\sigma_i)} \\
&= \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{16}\log_2(16) \\
&\quad + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{4}\log_2(4) \\
&= 3 \text{ bits}
\end{aligned}
$$

Example

A source emits one of four symbols s_0, s_1, s_2, and s_3 with probabilities 1/3, 1/6, 1/4, and 1/4, respectively. The successive symbols emitted by the source are statistically independent. Calculate the entropy of the source.

$$
\begin{aligned}
H(S) &= p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) + p_3 \log_2\left(\frac{1}{p_3}\right) \\
&= \frac{1}{3}\log_2(3) + \frac{1}{6}\log_2(6) + \frac{1}{4}\log_2(4) + \frac{1}{4}\log_2(4) \\
&= 0.528 + 0.431 + 0.5 + 0.5 \\
&= 1.959 \text{ bits}
\end{aligned}
$$