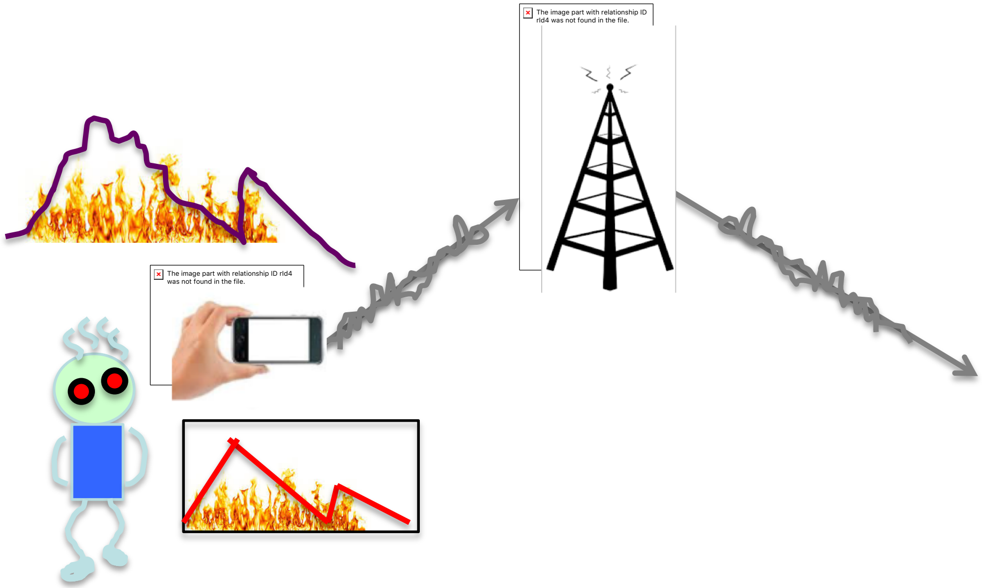
Re-cap

- Information theory & security
- Information communication system

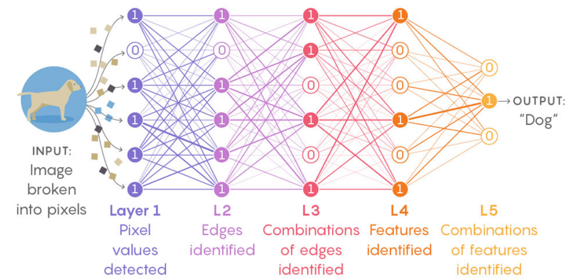


- How much “information” must be sent?

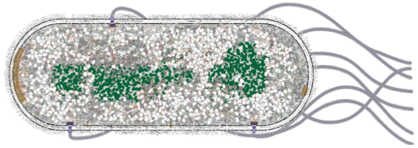
- How fast (rate) information can be transmitted so that Bob can recover

Learning From Experience

Deep neural networks learn by adjusting the strengths of their connections to better convey input signals through multiple layers to neurons associated with the right general concepts.

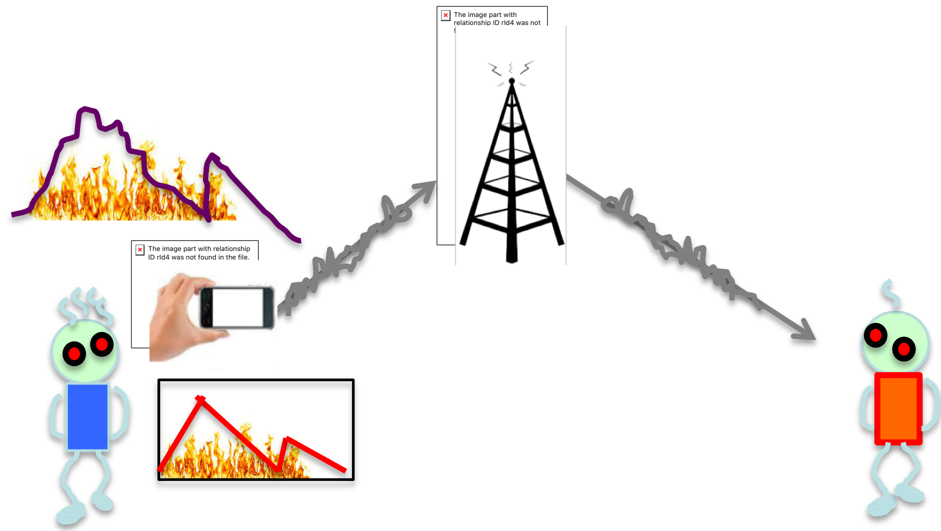


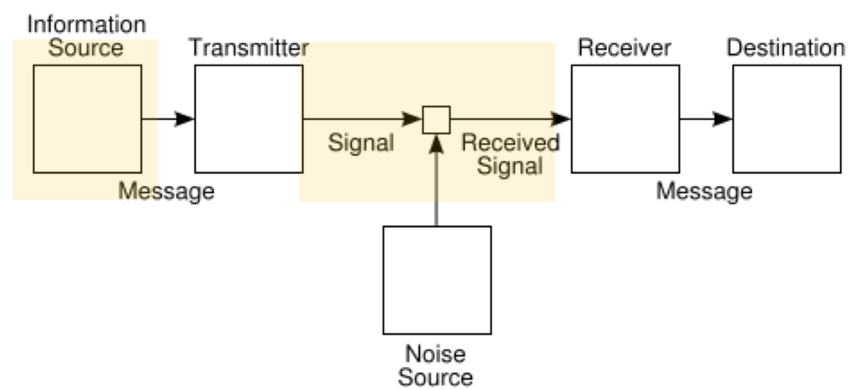
When data is fed into a network, each artificial neuron that fires (labeled “1”) transmits signals to certain neurons in the next layer, which are likely to fire if multiple signals are received. The process filters out noise and retains only the most relevant features.



Measuring information

- Measures for:
- Information source
  - Model
  - Entropy
- Channel
  - Relation of input and output
  - Joint entropy, mutual information



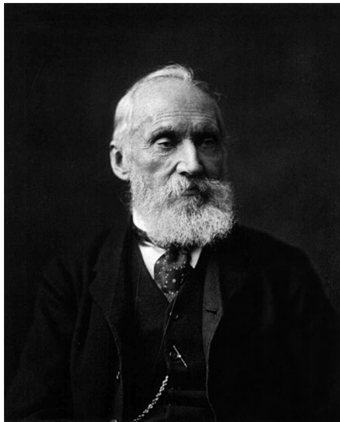


Measure of Information

- Centerpiece of Information Theory is the measure of information: entropy
  - Unit: bit

- “If you cannot measure it you cannot improve it.”

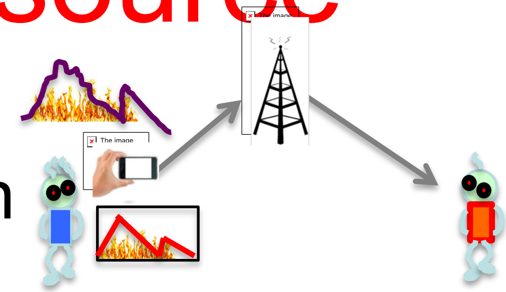
Lord Kelvin



Measuring information

Entropy of Information source

- Reconstruction without noise
  - A measure that matches our intuition



```
101010101010101
010101010101010
101010101010101
010101010101010
101010101010101
010101010101010
10101010101010
101010101010101
010101010101010
101010101010101
```

```
101010100101001
010001010111010
101000101010001
010101001001111
011100010101010
001001010100001
010011111000100
010000100001111
101001001010101
101001010001010
```

```
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
000000000000000
```

- Which one needs more “information” transfer?

Measuring information

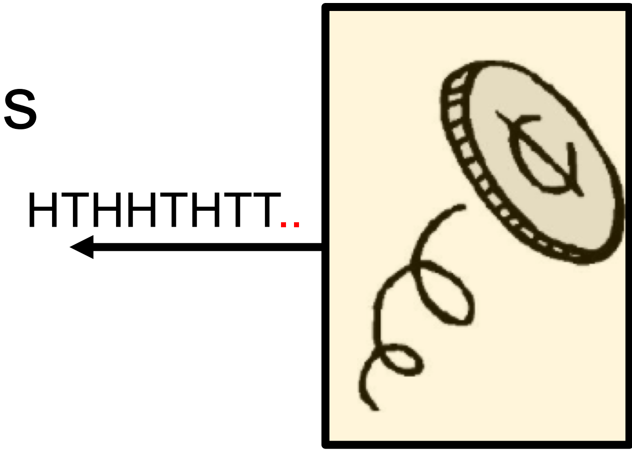
Information source

- Which of the following gives more information?

- Knowing the result of,
  1. One candidate election
  2. Three candidate election
- Information is related to uncertainty of outcome
  - "Information" removes uncertainly

A simple information source

- A coin toss machine
  - Knowing one of two alternatives



- Knowing the result when there are multiple alternatives

Guessing game

- *I choose a student* in the class: you find them

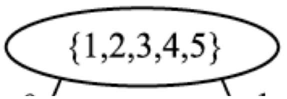
- How can you find the missing information?

- Ask questions
- Need to specify the type of questions:
  - Type of questions: Yes/No answer

- How many questions are needed to find the student?

Measuring information *is related to*
Number of questions needed to remove uncertainty

- Information can be quantified:
  - measured much the same way as we measure distance, time, mass, etc.



- In the guessing game:
  - the smallest number of questior a measure of 'information'
    - The shortest string of 0,1

- If questions are not carefully chosen, the number of questions will be larger than what is really needed.

How many question?

- For a class of N students, how many questions do we need*, assuming the best strategy?*

- $\log_2 N$

Logarithm (refresher)

- Logarithm is the inverse function of exponentiation

- Exponentiation: $2^3 = 8$
- Logarithm: $\log_2 8 = 3$

- $\log_2 4 = ?$
- $\log_2 32 = ?$
- $\text{Log}_2 17 = ?$

Hartley's information (1928)

- A finite set A with uniform distribution:
  - Each element has the same probability
- The average information learnt after knowing the outcome of selection:

$$H(A) := \log_b(|A|)$$

"A quantitative measure of 'information' is developed which is based on physical as contrasted by psychological considerations."

1888-1970 *Ralph Hartley*

Measure of Information: outcomes are not equiprobable

In most cases outcomes are not equi-probable.

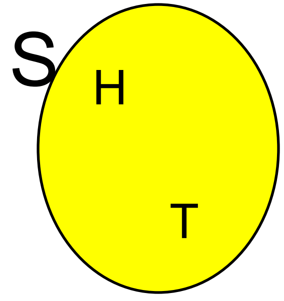
- Example: Election in the US:
  - Hillary Clinton, Donald Trump, Jill Stein

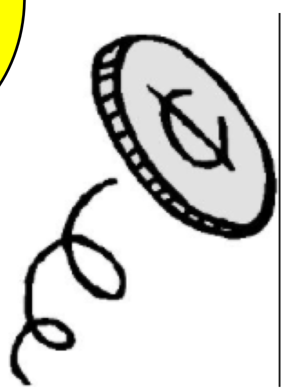
- How much information we have if we are given the election result?

- More “uncertain” $\leftarrow\rightarrow$ less probable $\leftarrow\rightarrow$ more information!

Probability *(refresher)*

- An experiment with uncertain outcome – possible outcomes sample space (discrete)
- A point in sample space has probability p_i
- $p_i \geq 0, \;\; \Sigma_i \, p_i = 1$

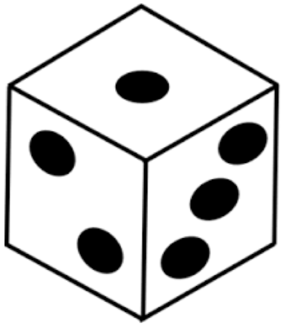




- Example:
- Outcome of a coin toss:
- S= {H,T}, $p_H = p(H) = 1/3$, $p_T = p(T) = 2/3$

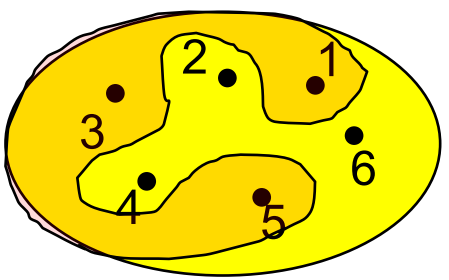
Probability *(refresher)*

- Experiment with uncertain outcome: Roll a die



- Sample space: possible outcome
- S={1,2,3,4,5,6}

- Probabilities
- Fair die : p(1)=p(2)…. =p(6)=1/6

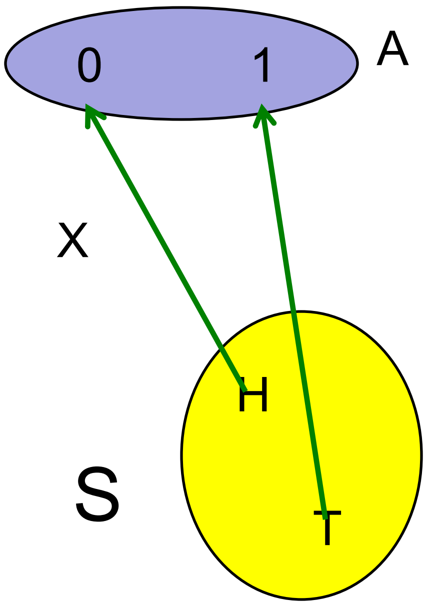


- Event : a set of points in sample
- Event E : outcome is odd

- p(E) = p(1)+p(3)+p(5)=1/2

Random Variables *(refresher)*

A random variable X is a function:



Domain: probability space
Range: a set A
- (need certain properties)

• X takes value in A, each with the associated probability in S

•Example:
•X maps coin tosses to {0,1}
- Defined over S
- A= {0,1}

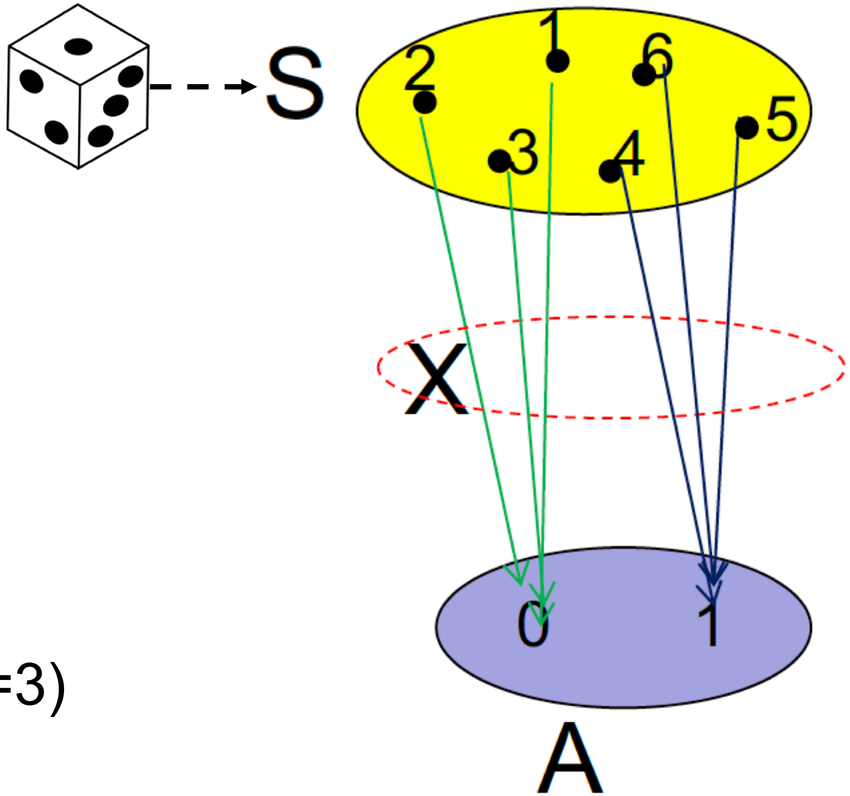
p(X=0)= p(H)=1/3
p(X=1)= p(T) = 2/3

Random Variables (refresher)

- Sample space: roll a die
- Biased die:
  p(1)=p(2)= p(3) =1/7
  p(4)=p(5)=p(6) =4/21

- Random variable X:
- X=0, S={1,2,3}
- X=1, S={4,5,6}

- p(X=0) =p(S=1) + p(S=2) + p(S=3)
  =1/7+1/7+1/7=9/21
- p(X=1) =p(S=4) + p(S=5) + p(S=6)
  =4/21+4/21+4/21=12/21



Random Variables *(refresher)*

- If the range of a random variable is numeric, we can define average (Expected value) of the variable:

  $A = \{a_1, a_2 \ldots a_m\}$ Expected value= $\Sigma_i \, a_i \, p(a_i)$

- Example: Sample space S={1,2,3,4,5,6},
- $p(1) = p(2) = p(3) = p(4) = p(5) = p(6)$

- Random variable Y: **Y= S mod 3** **$Y \in \{0, 1, 2\}$**

- Expected value of Y

  $1/3 \times 0 + 1/3 \times 1 + 1/3 \times 2 = 1$

Measure of Information: outcomes are not equiprobable

- A set of possible values- each with a different probability
- Example: Election in USA:
  - Donald Trump, Hillary Clinton, Jill Stein

- “Information” received from each outcome is different.

- Which statement has more “information”?
- “*Trump won the election*”
- “*Stein won the election*”

- How much information we receive from knowing each result?

Shannon Entropy

- X sample space
  $p(x_i)$ is probability of x_i
- Information gained by knowing the outcome is x_i is,

$$-\log_2 p(x_i) \text{ bit}$$

x_1—$p(x_1) \rightarrow -\log_2 p(x_i)$ bit

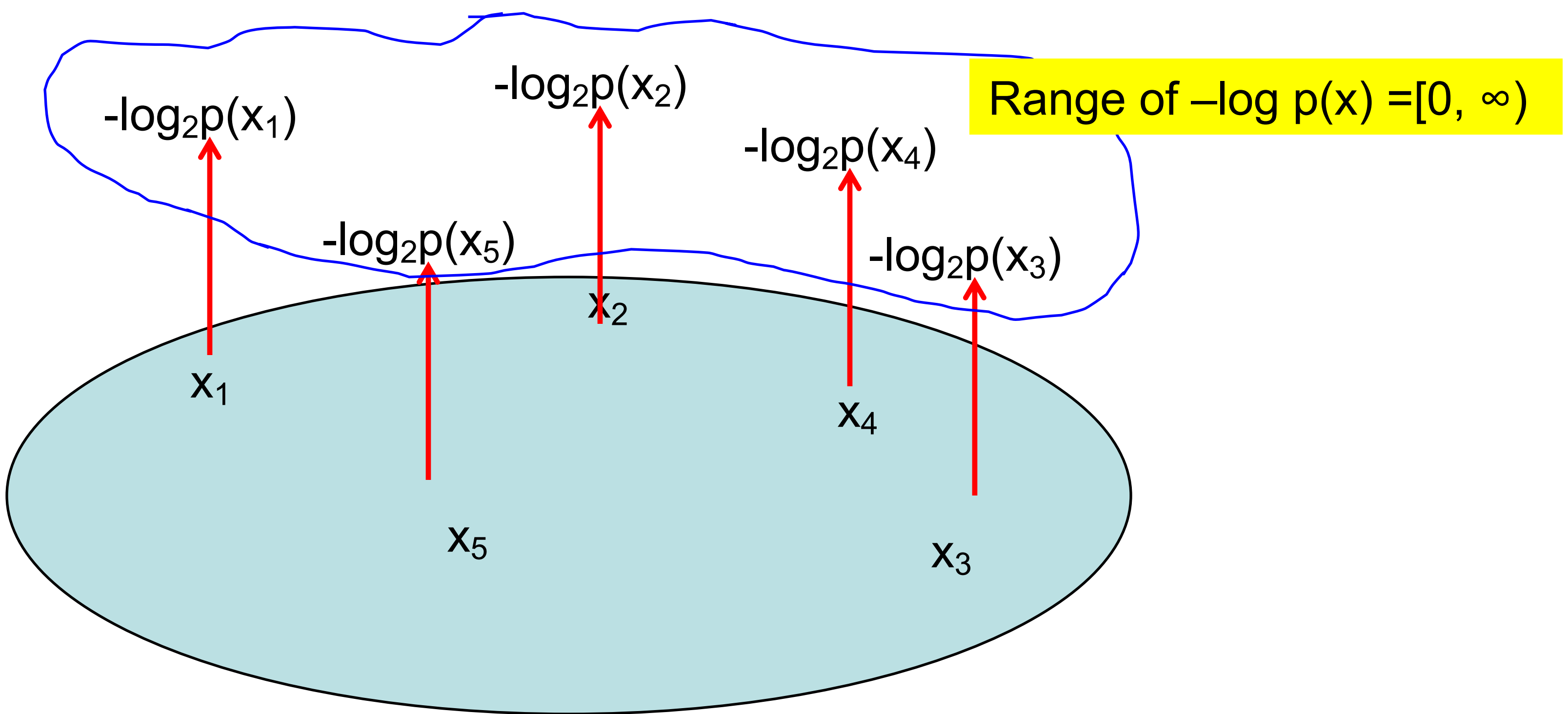
x_2—$p(x_2) \rightarrow -\log_2 p(x_2)$ bit

…

- Example:
- p(Clinton)=p(Trump) =0.49
- p(Stein)=0.02
- $-\log_2(0.49) = 1.029$ bit $-\log_2(0.02) = 5.64$ bit

Entropy

- Entropy is the average (expected) information from all outcomes:
- $H(X) = -\Sigma_i\, p(x_i) \log_2 p(x_i)$ bits



Example

- Tossing a coin:
- p_0=1/2, p_1=1/2

- $-\log_2 (1/2) = 1$ bit
- $-\log_2 (1/2) = 1$ bit
- Average information= $-(p_1 \log_2 p_1 + p_2 \log_2 p_2)$

- $(1/2)\, 1 + (1/2)\, 1 = 1$ bits

Example

- Tossing a coin:
- $p_0=1/3, \quad p_1=2/3$

- $-\log_2 (1/3) \ = \ 1.58$ bit
- $-\log_2 (2/3) \ =0.58$ bit
- Average information= $-(p_1\log_2 p_1 + p_2\log_2 p_2)$

- $(1/3)\ 1.58 + (2/3) 0.58 = 0.91$ bits

"Bit" as measure of information

- Different from bit as "binary digit"

- Equi-probable {head, tail }→ 1 bit of information

- "binary digit": Use ASCII characters to encode:
  - head : 68, 65, 61, 64
    - 01101000, 01100101, 01100001, 01100100
      32 binary digits
  - tail : 74, 61, 69, 6c

“Bit” as measure of information

- Different from bit as “binary digit”
  – unit of storage/transmission

- Store/reproduce an outcome {H, T}:
  – Use ASCII to encode → 8 bits
  – Information(equal prob) → 1 bit information

Example

- Two unbiased coins are flipped
- X is a random variable the shows the number of heads
- How much information do we learn when the outcome is 2 heads?

- X= {0,1,2}
- p(X=0) = 1/4, p(X=1) = 1/2 , p(X=2) = 1/4,

Less likely outcomes

$-\log_2 p(X=0) = -\log_2 p(X=2) = 2$ bits

$-\log_2 p(X=1) = 1$ bit

$H(X) = -\sum_i p(x_i) \log_2 p(x_i) = 1.5$ bits

What would be the questioning strategy?