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?

Less likely outcomes

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

$-\log_2 p(X=0) = -\log_2 p(X=2) = 2$ bits
$-\log_2 p(X=1) = 1$ bit

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

What would be the questioning strategy?

Questioning strategy

1. Is there one head?
2. Are there two heads?

   – On average in half of the cases the first question is enough.
   – If we need the second question then its answer determines whether there are two heads or none.

➔ in half of the cases we need 1 and the other half 2 questions

➔ Average number of questions: 1.5

Average number of questions is equal to the source entropy.

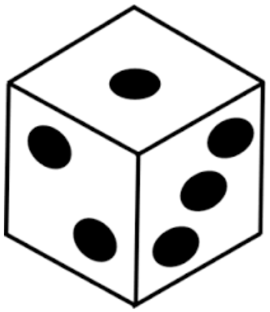
Shannon's entropy & Hartley's information

- If X has uniform distribution then Shannon entropy equals Hartley's measure of information (in base 2).

- $p(x_i) = 1/|X|$

$$
\begin{aligned}
H(X) &= -\Sigma_i \, p(x_i) \log_2 p(x_i) \\
&= -|X| \times 1/|X| \times \log_2 p(x_i) \\
&= -\log_2 p(x_i) = \log_2 |X|
\end{aligned}
$$

Example



- Entropy of an unbiased dice
  - How much information you will have on average, if you know the outcome?

$$-\ \frac{1}{6}\log_2(1/6) - \frac{1}{6}\log_2(1/6) - \frac{1}{6}\log_2(1/6) - \frac{1}{6}\log_2(1/6) - \frac{1}{6}\log_2(1/6) - \frac{1}{6}\log_2(1/6)) =$$

$$-\log_2(1/6) = 2.58 \text{ bits}$$

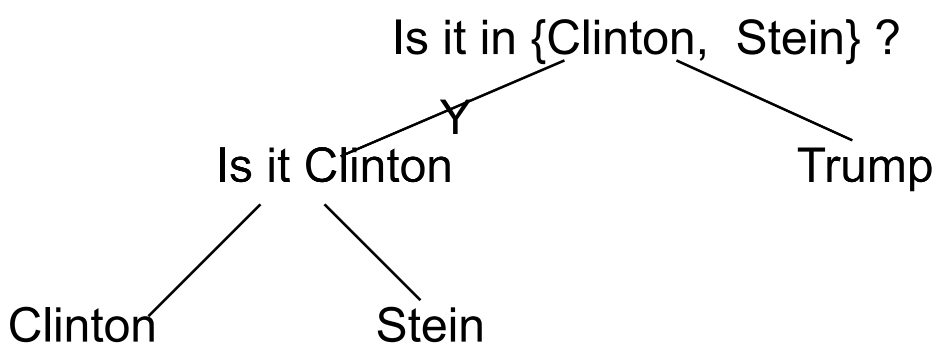
Hartley's information = $\log_2(6)$ =2.58 bits

Summary

- Probabilistic view of the world
  - Information is the complement of uncertainty
  - Information is the number of bits to “reconstruct” the signal
- Hartley’s measure of information
  - Equally likely outcomes
- Shannon’s measure
  - Non-uniform distribution

Election in the US

- Example: Election in the USA:
  - Donald Trump (0.49), Hillary Clinton (0.49), Jill Stein (0.02)
  - $H(X) = -0.49 \log_2(0.49) + 0.02 \log_2(0.02) = 1.21$ bits
- Amount of information in terms of number of questions:
  - Questioning strategy?
- Expected number of questions = 0.49 x1 + 0.51x 2=1.51



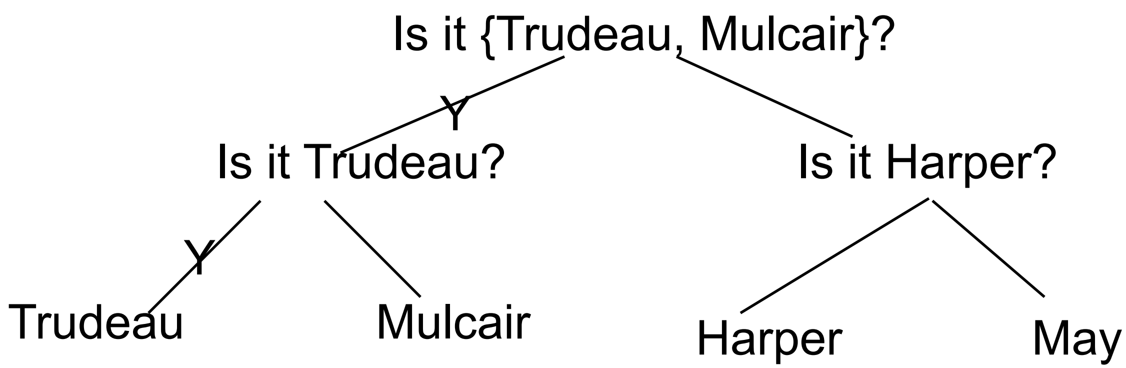
$H(X) \leq$ Num of questions $\leq H(X)+1$

Election in Canada

- Example: Election in Canada:
  - Tom Mulcair (0.3) , Justin Trudeau (0.3) , Stephen Harper(0.3) , Elizabeth May (0.1)

$$H(X) = -(0.3 \log_2 0.3) \times 3 - 0.1 \log_2 0.1)=1.9 \text{ bits}$$

- Questioning strategy
- Expected number of questions = 2



$H(X) \leq$ Num of questions $\leq H(X)+1$

Entropy & “number of questions”

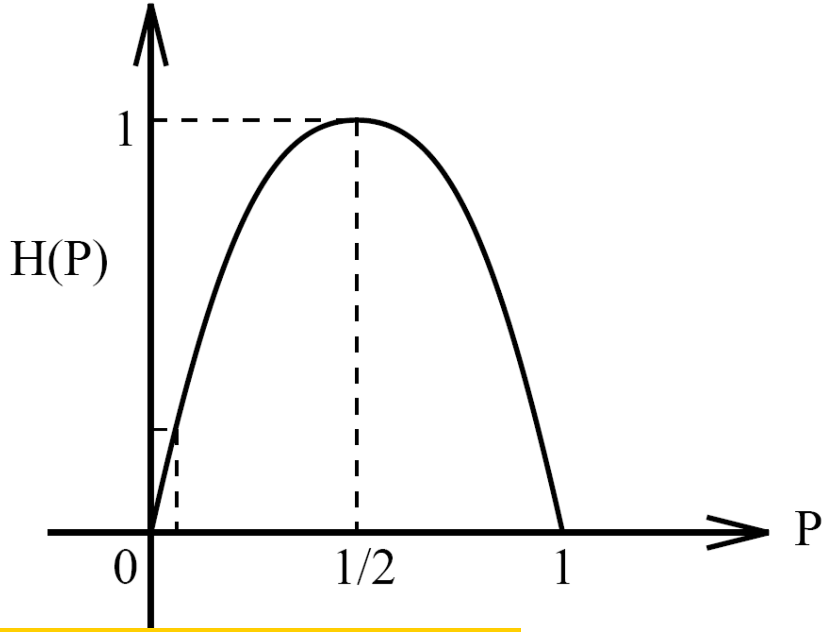
- H(X): Shannon entropy of a random variable X defined over a sample space $\mathcal{X}$

- Num : Average (Expected) number of questions to find an element of $\mathcal{X}$

- $H(X) \leq \text{Num} \leq H(X)+1$

Entropy of *Binary Random Variable*

- $\mathcal{X} = \{0, 1\}$
- $p(X = 1) = p$, $p(X = 0) = 1 - p$.
  $H(X) = -p \log p - (1 - p) \log(1 - p)$

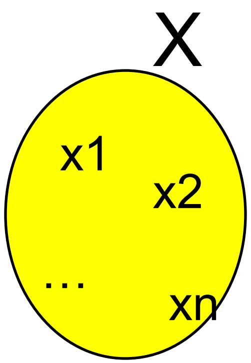


- $p = 1/2 \rightarrow H(X) = 1$ bit.
- $p = 0$ or $1 \rightarrow H(X) = 0$ bit

Entropy of unbiased coin is the highest.

Maximum Entropy

- An experiment with possible outcomes – $\mathcal{X}=\{x_1,x_2\ldots\ldots x_n\}$
- probability $p_i = p(X=x_i)$, $p_i \geq 0$, $\Sigma_i\, p_i = 1$



- Which distribution maximizes H(X)?
- Intuitively, uniform distribution: $p_i = 1/|X| = 1/n$
- This can be proved mathematically for
  $H(X) = -\Sigma p_i \log_2 p_i$

Choosing entropy function *(Shannon)*

- An information measure should satisfy certain properties.
- Non-negative
- If we have more choices, information that we receive would be more.
- An outcome that is less likely gives us more information
- Information received from two "independent sources" should be added
- Shannon proved that using above axioms, and some mathematical properties for information function, then $H(X) = - K \Sigma p_i \log_2 p_i$. is the only function

Plan

- Entropy is a measure of
  - Information (for reconstruction)
  - Uncertainty (how surprising)
- Entropy is used for measuring password strength
- Probabilistic modeling of information source
- Encoding source output using binary digits
  - Efficient codes
- Expected length of the best code

Password “entropy”

- Entropy is used as a measure of password strength
  - “uncertainty” about password
  - How many guesses/tries to find password

Password "entropy"

- Password entropy depends on the set of characters.

- 5 character passwords:
  - A= {a,b….z} → $5 \log_2 26 \sim 23.5$ bit
  - A= {a,b,…z, A, B, …Z} → $5 \log_2 52 \sim 28.5$ bit
  - A= {a,b,…z, A, B, …Z, 0,1..9} → $5 \log_2 62 \sim 29.7$ bit

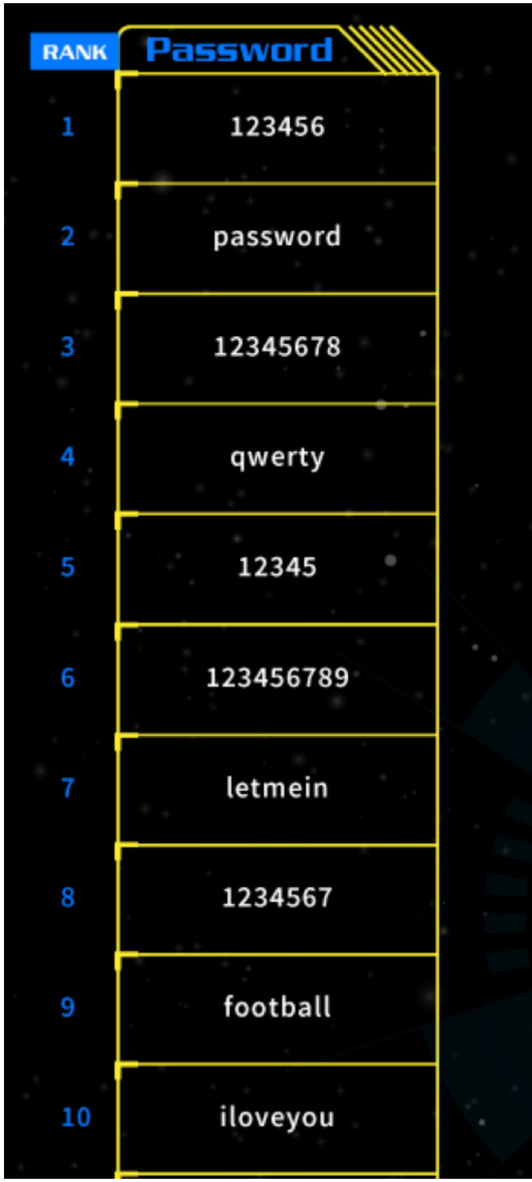
Password "entropy"

- Entropy of 6 digits passwords.
  - Size of the set of all passwords: 10^6
  - Entropy= $6 \times \log_2(10) = 6 \times 3.32 \sim 19.93$ bit
- Entropy of 6 character passwords, English letters and digits
  - Size of the set of all passwords: 36^6
  - Entropy= $6 \times \log_2(36) = 6 \times 5.17 \sim 31$ bit
- The numbers are upper bounds.
  - We assumed all passwords are chosen with the same probability
- What is the real entropy?

Password “entropy”

- “The password solutions company SplashData compiled a list of most common passwords based on data of five million passwords that were leaked by hackers in 2017. ”

- Passwords are not uniformly chosen.



https://interestingengineering.com/the-25-most-common-passwords-of-2017

Password "entropy"

- Need to know probability distribution of passwords
  - Need a large set of passwords

- Using data to learn how passwords are generated

In practice

- NIST Digital Identity guideline – June 2017
  - https://pages.nist.gov/800-63-3/sp800-63b.html
- Required entropy and size of character set depends on Authenticator Assurance Level
  - 20, 64 bit entropy

In practice

- Password strength meters use resistance against attacks:
  - Brute force attack
  - Dictionary attack
  - …

- Learning from corpus of published passwords

Summary

- Shannon’s entropy is the average amount of information per source output.

- Number of binary questions, using best questioning strategy is bounded by entropy

- Entropy for measuring password strength
  – unpredictability