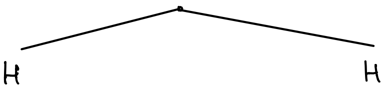
coin flip




$P(H) = 1$

Flip!

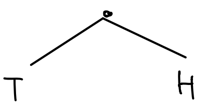
How much information did you get out of it? 0 "bits"

We will define a bit of information as:

"How many Yes/No questions you need to ask to determine the outcome?"

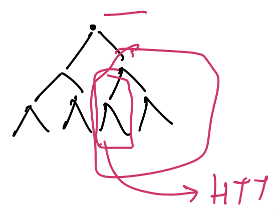
$P(H) = 0.5$

"1 bit"




Now, let's do 3 coin flips

3 - bits

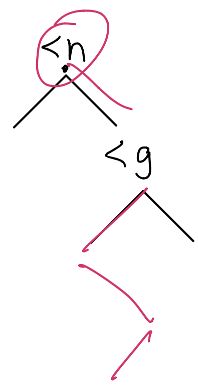



Pick a random letter from A – Z

What question should you ask?

A? B? C? $\rightarrow$ ~~26~~ 25 questions

a b c d e f g ~~h i j k l m n~~ o p q r s t u v w x y z




no. of questions $= \log_2(26)$

$= 4.7$ bits

Shannon

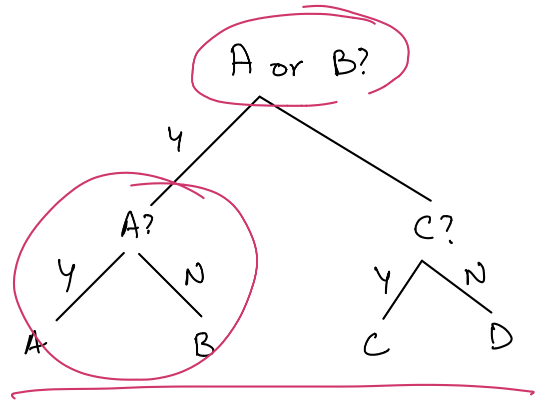
Let's consider just ABCD.

We have a sequence AABBCCDD

What's $P(A) = 2/8 = 1/4$

# of questions to find out it's an A $= 2 = \log_2(4)$

Same for B, C and D




Total "Expected value" for number of questions

$$= \sum_{\ell} P(L=\ell) \cdot Q$$

# of questions for this letter

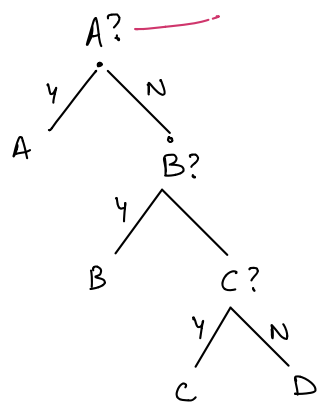
prob. of this letter

Expected number of questions $= \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2$

$= 2$ questions (on average)

notice this is $\log_2(4)$

Let's change the sequence to AAAABBCD




What's $P(A) = 4/8 = 1/2$

# of questions to find out it's an A $= 1$

$\boxed{\text{Expected number of questions:}}$

$$\rightarrow \underbrace{\tfrac{1}{2}\cdot 1}_{A} + \underbrace{\tfrac{1}{4}\cdot 2}_{B} + \underbrace{\tfrac{1}{8}\cdot 3}_{C} + \underbrace{\tfrac{1}{8}\cdot 3}_{D}$$

$= 1.75$

Average number of questions is lower!

Entropy

AA BB CC DD

Higher entropy 2

Unordered

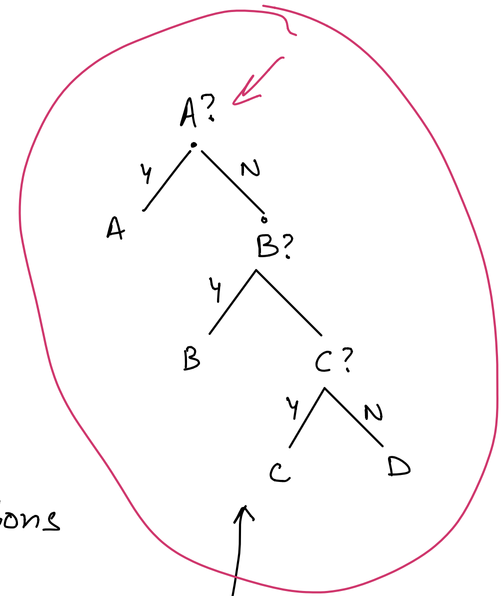
AAAA BB CD

lower entropy 1.75

Order

Points to note :

— Have to ask more questions to "gain knowledge" if entropy is higher




— If a machine is trying to learn something, we want to minimize the entropy so that it "learns" with fewer questions

— When transmitting information, we want to minimize entropy (# of questions asked)

That is why Huffman coding works

Total "Expected value" for number of questions

$$= \sum_{\ell} P(L=\ell) \cdot Q$$

# of questions for this letter (information)

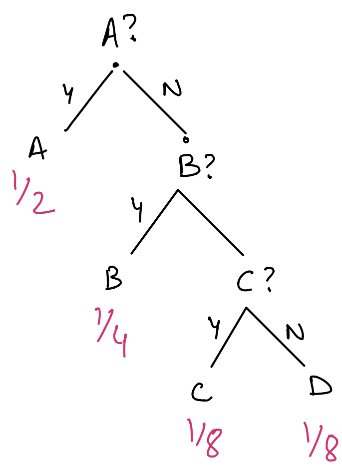
prob. of this letter

information

Q for A : 1 $= \underbrace{-\log_2 (0.5)}$ Prob

Q for B : 2 $= -\log_2 (0.25)$

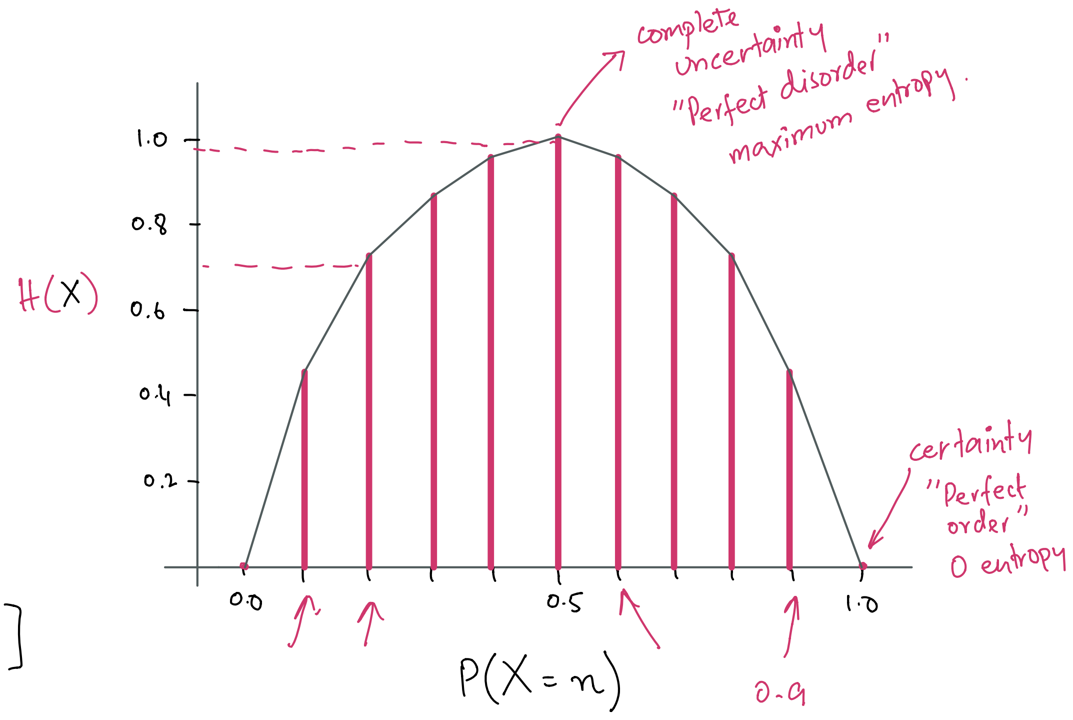
Q for C/D : 3 $= -\log_2 (0.125)$




→ Entropy $= H(X) = - \sum_i p_i \cdot \log_2 (p_i)$

So, entropy is "expected information"

Probability affecting entropy of a coin flip




$$H(x) = -\left[0.1 \underbrace{\log_2(0.1)}_{H} + 0.9 \underbrace{\log_2(0.9)}_{T}\right]$$

$$= \underline{0.46}$$

~~T = 1 H = 0 = 0~~

~~H = 1 T = 0 = 0~~