CHI - Basics of Information

- **Information**
  - ↳ data communicated or received that resolves uncertainty about a particular fact or circumstance

$$I(x_i) = \log_2\left(\frac{1}{p_i}\right)$$

$\log_2\left(\frac{1}{p_i}\right)$ ↝ uncertainty

- **Entropy** → avg info in piece of data
  - ↳ $H(X) = E(I(X))$

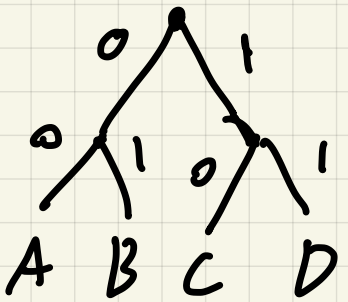
$$= \sum_{i=1}^{N} p_i \log\left(\frac{1}{p_i}\right)$$

- **Encoding** → mapping bits
  - map between bit strings and the set of possible data

e.g: A : 00
B : 01
C : 10
D : 11

Fixed length encoding → encoding where all choices are equally likely

e.g:



implies $H(x) = \log_2(N)$

Encoding Positive Integers

↳ $r = \sum_{i=0}^{N-1} 2^i b_i$

- smallest # = 0
- largest # = $2^N - 1$

Hexadecimal Notation

↳ base-16 representation

e.g: 0b0111010000 = 0x7D0

Encoding Signed Integers

↳ use two's complement encoding

- negative # have 1 → in most sig bit
- positive # have 0
- range: -2^{N-1} to $2^{N-1} - 1$
- $-A = \sim A + 1$

e.g: 1111 = -A → 0000 = ~A for 0

↳ 0001 = ~A + 1 → -1

Encoding Data (Variable length)

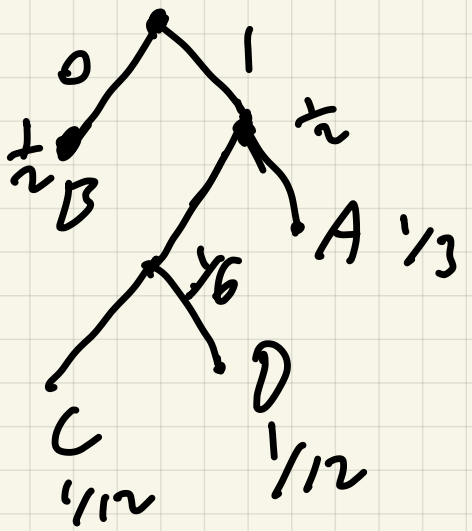
↳ higher prob. → shorter encod.

↳ lower prob. → longer encod.

Huffman Algo

↳ at each step in a tree, use the 2 symbols with highest prob.

e.g:



↳ to improve encoding efficiency) need to encode seq of choices not just choices themselves

Hamming Distance

↳ # of positions in which the corresponding digits of two encodings of the same length differ

↳ e.g 1 1 1 0 0 0 → Hamming dist = 2
1 1 0 0 1 0

Error Detection

↳ to detect E errors, you need a min Hamming dist of $E+1$

Error Correction

↳ to detect E errors, you need a min Hamming dist of $2E+1$