Entropy

Entropy

- (Information) Entropy aka *Shannon Entropy* is defined as: the average amount of information produced by a stochastic* source of data.
- What does that mean? Well, let's take an example a die. Each time we roll it, we get a number from 1 to 6.
- How much *information* is that? Well, the fact that we rolled a 6 tells us that we didn't roll a 1, 2, 3, 4, or 5. In particular, we have been given one of six equally-likely possible outcomes. There's no information value to the fact that we didn't roll 17—we knew that was impossible going in.

* *stochastic* is just a fancy word for randomized

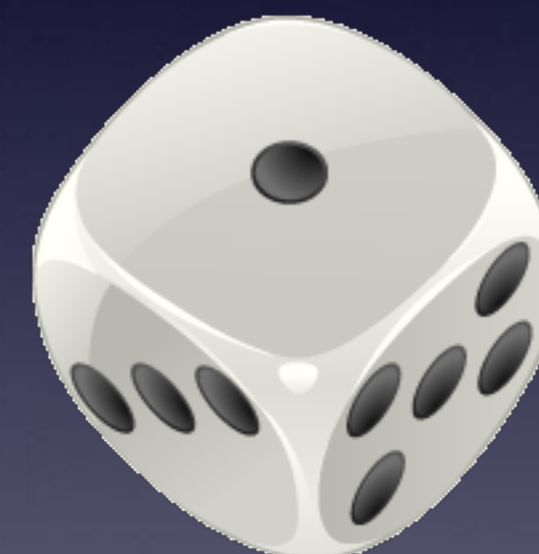
Entropy (continued)

- If we have *n* possible outcomes, each with probability p_i, then the average information (i.e. the entropy) is:

  $$h = -\sum_i p_i \log p_i.$$

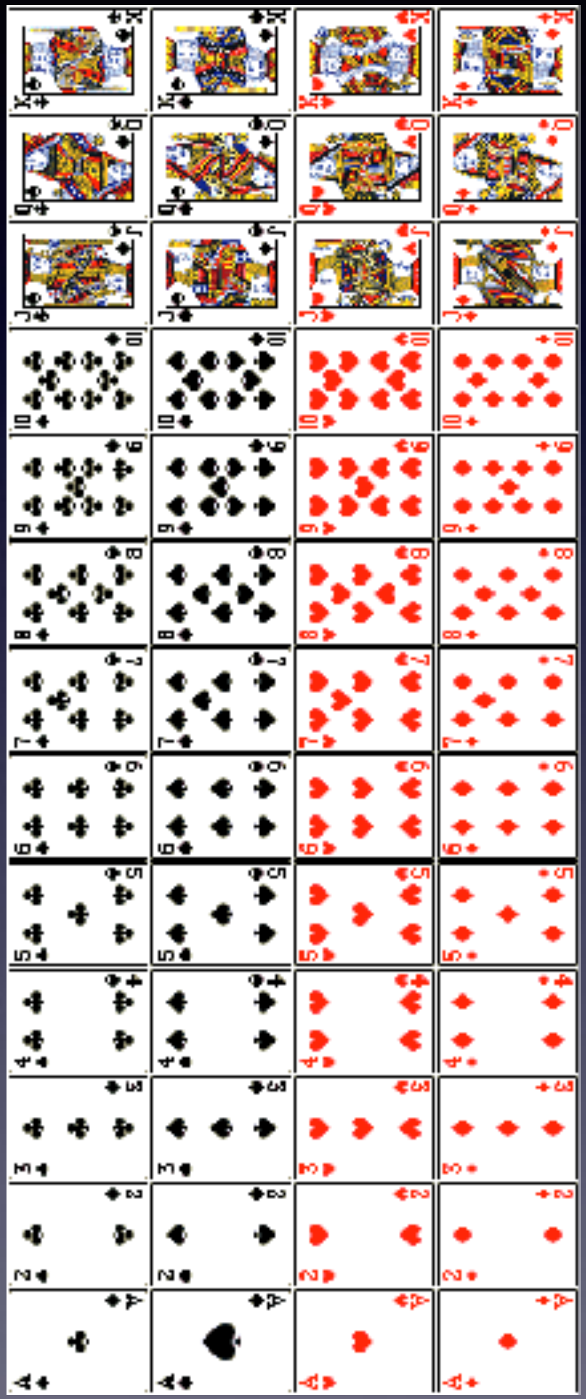
- We normally use log to the base 2 for entropy (which implies that the value can be expressed in *bits*).
- For the die with six possible outcomes,

  $h = 2.585$ bits where $2^h = 6$



Another example

- Drawing a card from a shuffled deck (52 cards):
  - h = 5.7
- You take the card but don't show me. I ask you several questions which you must answer truthfully:
- Is it a red or black card? Now, h = 4.7
- Is it a round suit or a pointy suit? h = 3.7
- Is it an "honor" (T, J, Q, K, or A)
  - If yes, then h = 2.32
  - If no, then h = 3
- Assume no, then is *k* an even number? h = 2
- Is *k/2* an even number? h = 1
- Is *k/4* an even number? h = 0
  - (Now I **know** which card it is because there is zero entropy—i.e. nothing is still unknown, no surprises to come).



What about a shuffled deck?

- The entropy (average information per card) is 5.7 so the total entropy in the deck is 51 * 5.7 = 290 [Actually this is a high estimate because as we go through the deck, cards aren't returned and so the entropy for each drawing is successively smaller].
- There are 52! (8 followed by 67 zeros) possible orderings of the deck and we don't know which one it's in until we've drawn 51 cards (we get the last by process of elimination). So another way to calculate the total entropy of the deck is simply to calculate:

  $lg\ (52!) = 225.6$

- Suppose we were trying to sort the deck so that it was in some pre-defined order? Let's say each operation we perform (like compare + swap) removes 1 bit of entropy.
- **We would need *at least* 226 operations to sort the deck.** We cannot do any better than that, however smart we are!

Generalizing our ideal sort

- We showed that to sort a deck of cards, we'll need at least 226 operations. But what about a list of *n* elements?
- What we did for the deck of cards was to calculate 52! and take its log to base 2.
- What about *n!* ? Is there a way we can approximate this so that we don't have to multiply it out (hard to do when we don't know the actual number!)
- It's called Stirling's Approximation:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

- What we actually want is the *lg* ($\log_2$):
  - $\sim n(\lg n - \lg e) + 1/2 \lg (2\pi n)$
  - $\sim n(\lg n - 1.44)$ [as *n* increases, the importance of the 2nd/3rd terms reduces]
- **This expression represents the theoretical minimum number of operations to sort our list of length *n*.**

Because of entropy, the best we can do when sorting a list of *n* elements is *n lg n* operations.