Quantum Information Theory

What is the information content of a state?

- Information is "what we can learn" (Claude Shannon)

  - Suppose we flip a coin. What do we learn about odds of H vs T?
    - If we flip a fair coin, we don't get a lot of information for each H or T. We know $P(H) \approx P(T) \approx \frac{1}{2}$ reasonably quickly, but it takes a while to nail down.
    - If $P(H) \gg P(T)$, each H tells us nearly nothing, but each T tells us a lot about $P(T)$
    - Information is about classical ignorance

  - Quantitatively, suppose we flip N coins. We expect $P(H)N \equiv pN$ heads and $(1-p)N$ tails.
    - The number of arrangements is (Stirling's Approx)

$$\# = \frac{N!}{(pN)!\,((1-p)N)!} \approx \frac{N^N}{(pN)^{pN}[(1-p)N]^{(1-p)N}} = \frac{1}{p^{pN}(1-p)^{(1-p)N}}$$

    - We define Shannon entropy S such that $\# = 2^{NS}$ so

$$S = -p\log_2 p - (1-p)\log_2(1-p)$$

This is like "how much" of a coin each flip contributes $\Leftarrow$ there are 2^S alternatives per flip

    - If there are multiple options w/probabilities P_i each, then

$$S = -\sum_i P_i \log_2 P_i$$

    - The entropy is the average of the information content of each flip value $i \equiv -\log_2 P$ in Shannon bits

- Alternately, we can think in terms of data compression
  + Consider an alphabet w/ 4 letters a, b, c, d
  + To encode these in binary, we can use
    $a = 00, \quad b = 01, \quad c = 10, \quad d = 11$
  + However, this is inefficient if some letters are more common in the language.
  + Suppose the probabilities for a given letter to take each value is
    $P(a) = 1/2, \quad P(b) = 1/4, \quad P(c) = P(d) = 1/8$
    The (most) efficient encoding is
    $a = 0, \quad b = 10, \quad c = 110, \quad d = 111$
    with the rule that 0 means "next letter" and max length = 3
  + Then the length of each letter is $i = -\log_2 P$
    and the Shannon entropy is the average letter length in the language

- Notes:
  + The ~~definition~~ definition for Shannon entropy is very similar to the Gibbs definition of the thermodynamic entropy (in terms of microstate probability).
  + There are many extensions in classical information theory, but we'll focus on the quantum theory.

– Quantum Entropy

- Quantum information is also what we can learn
  + If we have a pure state (and know what it is), you know all there is to know about the system even if you can't predict a specific measurement
  + In a mixed state, we can generalize the Shannon entropy to
    $$S = -\sum_i P_i \log_2 P_i$$
    where the P_i are the evalues of the density matrix. This is the

+ We can write this in terms of the density matrix

$$S = -\sum_i \langle e_i | \rho \log_2(\rho) | e_i \rangle = -tr(\rho \log_2(\rho))$$

where the sum is over the e'basis of ρ.
This is the <u>von Neumann entropy</u>

+ <u>von Neumann entropy of a pure state is $S=0$</u> ($P_1=1, P_{i\neq 1}=0$)
When a mixed state arises as part of an entangled state, von Neumann entropy may be called <u>entanglement entropy</u>

∘ Properties of von Neumann entropy (mostly presented w/o proof)

+ If the dimension of Hilbert space is N, entropy has max. value $S \leq \log_2(N)$, with equality only when the state is maximally mixed — all N eigenvalues of ρ are $P = 1/N$.

+ Suppose system 1 is in a mixed state, and we purify it by entangling it with system 2 (systems 1+2 together are in a pure state). Then $S(\rho_1) = S(\rho_2)$ for the entropies if we partial trace either system.
This is true even if systems 1+2 have different Hilbert space dimensions.

+ The entropy is <u>concave</u> (2^{nd} der matrix negative for normal functions)

$$\sum_i x_i S(\rho_i) \leq S(\textstyle\sum_i x_i \rho_i) \quad \text{where } x_i \geq 0 \text{ and } \sum_i x_i = 1$$

That is, mixing density matrices increases entropy.
(See reading for proofs)

+ Suppose you have a quantum system with 2 pieces.
If the overall density matrix is ρ,

$$S(\rho) \geq |S(\rho_1) - S(\rho_2)| \quad \text{\underline{Araki-Lieb inequality}}$$

and

$$S(\rho) \leq S(\rho_1) + S(\rho_2) \quad \text{\underline{subadditivity of entropy}}$$

where $\rho_1 + \rho_2$ are the reduced density operators

+ If a system has 3 parts 1, 2, + 3, Then

$$S(\rho_{1,2}) + S(\rho_{2,3}) \geq S(\rho_2) + S(\rho)$$

where ρ = overall density matrix

$\rho_{1,2}$ = reduced by partial trace over 3

$\rho_{2,3}$ = reduced by partial trace over 1

ρ_2 = reduced by partial trace over 1 + 3

This is strong subadditivity

- Other "entropies"

+ Many of these properties are proven using the relative entropy

$$S(\rho\|\sigma) = tr\left[\rho(\log_2\rho - \log_2\sigma)\right]$$

where ρ = density operator of the system and σ is an arbitrary reference density operator.

The key property is that $S(\rho\|\sigma) \geq 0$.

+ The Rényi entropy is defined for a number α as

$$S_\alpha(\rho) \equiv \frac{1}{1-\alpha}\log_2(tr\,\rho^\alpha)$$

This becomes von Neumann entropy in the limit $\alpha \to 1$

+ This is only a very short introduction and there is a lot more you can learn!