von Neumann Entropy

The Shannon entropy measures the uncertainty in a probability distribution

$$H(p) = \sum_i p_i \log \frac{1}{p_i} = -\sum p_i \log p_i \quad (\log \text{ base } 2 \Rightarrow \text{bits})$$

Ex: Fair coin has entropy $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = -\log\frac{1}{2} = \log 2 = 1$

This has many operational interpretations (e.g., compression — more later)

Quantum analog: von Neumann entropy, $S(\rho) = -\mathrm{tr}(\rho \log \rho)$ (Shannon ent. of eigenvalues)

Properties of entropy:

- $0 \le S(\rho) \le \log D$, where ρ is $D \times D$
- $S(\rho) = 0 \iff \rho$ is pure
- $S(\rho) = \log D \iff \rho$ is maximally mixed ($\rho = \frac{1}{D} I_D$)
- $S(\sum p_i \rho_i) \ge \sum p_i S(\rho_i)$ (concavity)
- $S(\sum p_i \rho_i) \le H(p) + \sum p_i S(\rho_i)$, with equality if the ρ_i have orthogonal support
- ρ_{AB} pure $\Rightarrow S(\rho_A) = S(\rho_B)$
- $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$
- Subadditivity: $S(\rho_{AB}) \le S(\rho_A) + S(\rho_B)$
- Araki–Lieb: $S(\rho_{AB}) \ge |S(\rho_A) - S(\rho_B)|$
- Strong subadditivity: $S(\rho_{ABC}) + S(\rho_B) \le S(\rho_{AB}) + S(\rho_{BC})$

Entropy can also be seen as a measure of the entanglement of a bipartite pure state $|\psi\rangle_{AB}$: $E(|\psi\rangle) = S(\mathrm{tr}_A |\psi\rangle\langle\psi|) = S(\mathrm{tr}_B |\psi\rangle\langle\psi|)$

$E(|\psi\rangle) = 0$ iff $|\psi\rangle$ is a product state
$E(|\beta_{00}\rangle) = 1$
$E((U \otimes V)|\psi\rangle) = 0$; more generally, LOCC-invariant

Using LOCC, for large n, with fidelity $\to 1$:

$$|\psi\rangle^{\otimes n} \longrightarrow |\beta_{00}\rangle^{\otimes nE(|\psi\rangle)} \quad \text{(entanglement concentration)}$$
$$|\beta_{00}\rangle^{\otimes nE(|\psi\rangle)} \longrightarrow |\psi\rangle^{\otimes n} \quad \text{(entanglement dilution)}$$

Since this is reversible (asymptotically), $|\beta_{00}\rangle$ can be seen as a standard unit of (pure state) entanglement, and $E(|\psi\rangle)$ is a unique measure of (pure state) entanglement.

Situation is more complicated for mixed states!

Compression

Suppose Alice wants to send messages to Bob as efficiently as possible.
If all messages are not equally likely, she can compress the input to send the likely messages more efficiently.

Ex: 4 messages: A (prob. $\frac{1}{2}$), B (prob. $\frac{1}{4}$), C (prob. $\frac{1}{8}$), D (prob. $\frac{1}{8}$)
Sending as is would take $\log 4 = 2$ bits
instead, encode $A \to 0$, $B \to 10$, $C \to 110$, $D \to 111$
expected # bits: $\frac{1}{2} + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = \frac{7}{4} < 2$
compare to entropy: $\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + 2 \cdot \frac{1}{8}\log 8 = \frac{7}{4}$
note: in encoded message, all symbols are 0 or 1 with equal probability

To make things nice in general, use <u>block coding</u>: collect many messages from same source before encoding

Assume source is i.i.d. Simple case: source outputs bits, 0 w.p. p, 1 w.p. 1-p.
With n samples, $\Pr(x_1 \dots x_n) = p^{\alpha n}(1-p)^{(1-\alpha)n}$ where α = fraction of 0s
Typically, $\alpha \approx p$. We say $x_1 \dots x_n$ is a <u>typical sequence</u> if this holds.

Probability of a sequence being typical: $\approx p^{pn}(1-p)^{(1-p)n}$
log of this: $pn\log p + (1-p)n\log(1-p) = -nH_2(p)$
where $H_2(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy

so $\Pr(\text{typical}) \approx 2^{-nH_2(p)} \implies$ # of typical sequences $\leq 2^{nH_2(p)}$
thus we can specify which typical sequence using $nH_2(p)$ bits
if the sequence is not typical, abort (failure prob. will be low) or send uncompressed (expected length will still be small)

To make this rigorous, need to quantify typicality: a sequence is ε-typical if its probability is between $2^{-n(H(p)+\varepsilon)}$ and $2^{-n(H(p)-\varepsilon)}$.
Can prove that for any $\varepsilon, \delta > 0$, for large enough n,
- $\Pr(\text{sequence is } \varepsilon\text{-typical}) \geq 1-\delta$
- $(1-\delta)2^{n(H(p)-\varepsilon)} \leq (\#\ \varepsilon\text{-typical seq.}) \leq 2^{n(H(p)+\varepsilon)}$

Call a compression scheme <u>reliable</u> if the probability of correct transmission $\to 1$ as $n \to \infty$. The <u>rate</u> of a scheme is R if it compresses n bits $\to nR$ bits.

<u>Shannon's noiseless coding theorem</u>: Consider an i.i.d. source with distribution p.
If $R > H(p)$ then there is a reliable scheme of rate R.
If $R < H(p)$ then no scheme of rate R is reliable.

Schumacher compression

What about quantum information?

An i.i.d. q. source is described by an ensemble where ρ_i occurs w.p. p_i
Described by density matrix $\rho = \sum p_i \rho_i$
Diagonalize: $\rho = \sum \lambda_i |\psi_i\rangle\langle\psi_i|$
Main observation: ρ is indistinguishable from source that sends $|\psi_i\rangle$ w.p. λ_i

Strategy: project onto the typical subspace, eigenvectors of $\rho^{\otimes n}$ corresponding to typical sequences of $|\psi_i\rangle$'s
Here an eigenvector is ε-typical if $2^{-n(S(\rho)+\varepsilon)} \leq \prod \lambda \leq 2^{-n(S(\rho)-\varepsilon)}$
Most of the state lives in the typical subspace, so this hardly changes the state
After projecting, we can describe the state with only $\approx n S(\rho)$ qubits

By analogy to Shannon:
- An encoding scheme is reliable if the fidelity of the output with the input $\xrightarrow{n\to\infty} 1$
- A scheme has rate R if it encodes n qubits into nR qubits

Theorem (Schumacher): Consider an i.i.d. q. source with avg. state ρ.
If $R > S(\rho)$ then there is a reliable scheme with rate R.
If $R < S(\rho)$ then any scheme of rate R is not reliable.

Entanglement concentration/dilution

Typical sequences can also be used to describe asymptotic entanglement transformations

Concentration: given $|\psi\rangle^{\otimes n}$, $|\psi\rangle = \sum_i \sqrt{p_i}\,|i\rangle_A|i\rangle_B$ (Schmidt decomp: $\sum_{ij} a_{ij}|i\rangle|j\rangle$, SVD: $a_{ij} = \sum_k u_{ik} d_{kk} v_{kj}$, Reduced state: $\sum_i p_i |i\rangle\langle i|$)

so $|\psi\rangle^{\otimes n} = \sum_{i_1,\ldots,i_n} \sqrt{p_{i_1}\cdots p_{i_n}}\,|i_1\ldots i_n\rangle|i_1\ldots i_n\rangle$

project onto typical subspace; typical sequences all have probabilities $\approx 2^{-nS(\rho)}$
however, they're not close enough in general for this state to be close to max. ent.
instead, also measure the type class (for bits, the Hamming weight)
now all strings have the same amplitude, and likely type classes have size almost $nS(\rho)$

Dilution: given $|\beta_{00}\rangle^{\otimes nR}$, produce $|\psi\rangle^{\otimes n}$

this can be done easily using Schumacher compression + teleportation
create $|\psi\rangle_{AC}^{\otimes n}$ locally
compress C part to $\approx nS(\rho)$ qubits
teleport these using $nS(\rho)$ ebits (and $2nS(\rho)$ bits)
decode the received qubits and uncompress