I. von Neumann entropy and its properties

The von Neumann entropy is just the Shannon entropy of the state's eigenvalues:

$$S(\rho) = -\sum_k \lambda_k \log \lambda_k = -\mathrm{Tr}[\rho \log \rho]$$

Properties:

1. $S(\rho) = 0$ iff ρ is a pure state
2. $S(U\rho U^\dagger) = S(\rho)$ for unitary U
3. $S(\rho) \leq \log|\mathrm{supp}\,\rho|$
4. $S(\sum_k p_k \rho_k) \geqslant \sum_k p_k S(\rho_k)$
5. $S(\sum_k P_k \rho P_k) \geqslant S(\rho)$ for any complete set of projectors P_k

Properties 1 and 2 are clear by inspection. Just like the classical case, the rest follow from the positivity of the relative entropy, a statement known as Klein's inequality:

$$S(\rho\|\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)] \geqslant 0$$

with equality iff $\rho = \sigma$.

3. Define $\sigma = \mathbb{1}/d$ where $\mathbb{1}$ is identity operator on $\mathrm{supp}\,\rho$ and $d = |\mathrm{supp}\,\rho|$. Then $S(\rho\|\sigma) = \log d - S(\rho) \geqslant 0$

4. Let $\rho = \sum_k p_k \rho_k$. Then $\sum_k p_k S(\rho_k\|\rho) = S(\rho) - \sum_k p_k S(\rho_k) \geqslant 0$

5. Let $\bar\rho = \sum_k P_k \rho P_k$. Observe $[P_k, \bar\rho] = 0$. $S(\rho\|\bar\rho) = -S(\rho) - \sum_k \mathrm{Tr}[P_k(\rho \log\bar\rho)P_k]$

$$= S(\bar\rho) - S(\rho) \geqslant 0$$

II. Justifications / Operational Interpretations

A. Schumacher compression

Schumacher compression is the quantum analog of classical data compression / source coding, where the source produces quantum states instead of classical random variables.

Suppose a source produces states $|\varphi_k\rangle \in \mathbb{C}^d$ with probability p_k.
Can we compress n outputs of this source into a state space of fewer than d^n dimensions? Yes! The idea is to work in the eigenbasis of the average state $\rho = \sum_k p_k |\varphi_k\rangle\langle\varphi_k|$ and treat the problem as classical source coding.

Let $\rho = \sum_k p_k |\varphi_k\rangle\langle\varphi_k| = \sum_j \lambda_j |j\rangle\langle j|$

Then $\rho^{\otimes n} = \sum_{\vec{j}} \lambda_{\vec{j}} |\vec{j}\rangle\langle\vec{j}|$, where $|\vec{j}\rangle = |j_1\rangle|j_2\rangle \cdots |j_n\rangle$ and $\lambda_{\vec{j}} = \lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_n}$, has most of its support on the typical subspace T_n^ε, defined to be the space spanned by eigenvectors whose eigenvalues are typical.
Let P be the projector onto this subspace. Then we can define an encoder $\mathcal{E}$ by $\mathcal{E}(\rho^{\otimes n}) = VP\rho^{\otimes n}PV^\dagger + \rho_{\text{fail}}\left(1 - \mathrm{Tr}[P\rho^{\otimes n}]\right)$,
that is, the encoder measures $\rho^{\otimes n}$ with P and $\mathbb{1}-P$. If the input is typical, then the encoder uses the partial isometry V to transfer the "pruned" state to a state space of $2^{n[S(\rho)+\varepsilon]}$ dimensions. If the input isn't typical, it outputs the state ρ_{fail}, which we take to live in some other state space.

That way, the decoder can look at the output, decide if it is "good" or not, and proceed to apply $V^\dagger$ if it is. Then we have the following protocol:

$$\boxed{\text{source}} \xrightarrow{\rho^{\otimes n}} \boxed{\text{encoder}} \xrightarrow{\sigma} \boxed{\text{decoder}} \longrightarrow \tilde{\rho}$$

$$\text{dim} = d^n \qquad\qquad \text{dim} = 2^{n(S(\rho)+\varepsilon)}$$

How good an approximation to $\rho^{\otimes n}$ is the output $\tilde{\rho}$?

We can quantify this by the trace distance $\|\rho^{\otimes n} - \tilde{\rho}\| = \frac{1}{2}\mathrm{Tr}|\rho^{\otimes n} - \tilde{\rho}|$ (*)

The output is given by $\tilde{\rho} = P\rho^{\otimes n}P + (1-\mathrm{Tr}[P\rho^{\otimes n}])\,\rho_{fail}$, so

$$\frac{1}{2}\mathrm{Tr}\left|\rho^{\otimes n} - P\rho^{\otimes n}P - (1-\mathrm{Tr}[P\rho^{\otimes n}])\rho_{fail}\right|$$

$$= \frac{1}{2}\left[(1-\mathrm{Tr}[P\rho^{\otimes n}]) + \mathrm{Tr}\left|\rho^{\otimes n} - P\rho^{\otimes n}P\right|\right]$$ ρ_{fail} lives on an orthogonal space to $\rho^{\otimes n}$

$$= \frac{1}{2}\left[1-\mathrm{Tr}[P\rho^{\otimes n}] + \sum_{\vec{k}}\left|\lambda_{\vec{k}} - \lambda'_{\vec{k}}\right|\right] \qquad \lambda'_{\vec{k}} = \begin{cases} \lambda_{\vec{k}} & \text{if } \vec{k} \in T_n^\varepsilon \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{2}\left[1-\mathrm{Tr}[P\rho^{\otimes n}] + \sum_{\vec{k}\notin T_n^\varepsilon}\lambda_{\vec{k}}\right]$$

$$= 1-\mathrm{Tr}[P\rho^{\otimes n}]$$ but $\mathrm{Tr}[P\rho^{\otimes n}] > 1-\delta$

$$< \delta$$

(*) $|A| = \sqrt{A^\dagger A} = \sum_k |\lambda_k|$ if A Hermitian with eigenvalues λ_k

The trace distance is perhaps the best measure to use, because it can be shown that if $\|\rho - \sigma\| < \varepsilon$, then we can think of σ as being ρ with prob. $\geq 1-\varepsilon$. That is, no measurement we could perform would give different outputs assuming ρ rather than σ with total prob. $> \varepsilon$