always imagine that A is mixed because it was entangled with some other system B. All we need is to make that formal. One thing we note from the start is that this operation is certainly not unique since the system B can have any size. Thus, there is an infinite number of pure states which purify ρ_A. The simplest approach is then to consider B to be a copy of A. We then define the pure state

$$|\psi\rangle = \sum_a \sqrt{p_a} |a\rangle \otimes |a\rangle \tag{2.109}$$

Tracing over B we get

$$\mathrm{tr}_R \, |\psi\rangle\langle\psi| = \rho \tag{2.110}$$

Thus, $|\psi\rangle$ is a purified version of ρ, which lives in a doubled Hilbert space. Notice how the probabilities p_a appear naturally here as the Schmidt coefficients.

2.9 Entropy and mutual information

The concept of entropy plays a central role in classical and quantum information theory. In its simplest interpretation, entropy is a measure of the disorder (or mixedness) of a density matrix, a bit like the purity $\mathrm{tr}(\rho^2)$. But with entropy this disorder acquires a more informational sense. We will therefore start to associate entropy with questions like “how much information is stored in my system”. Also like the purity, entropy can be used to quantify the degree of correlation between systems. And that makes sense because correlation is a measure of information: when two systems are correlated we can ask questions such as “how much information about A is stored in B”. Unlike the purity, however, entropy will also serve to quantify correlations of mixed states, which is done using the concept of **mutual information**. We will also introduce another concept called **relative entropy** which plays the role of a “distance” between two density matrices. It turns out that the relative entropy is not only useful in itself, but it is also useful as a tool to prove certain mathematical identities.

In thermodynamics we like to associate entropy with a unique physical quantity. In quantum information theory that is not exactly the case. There is one entropy, called the **von Neumann entropy**, which does have a prominent role. However, there are also other entropy measures which are also of relevance. An important family of such functions are the so-called **Rényi entropies**, which contain the von Neumann entropy as a particular case. We will also discuss them a bit.

The von Neumann entropy

Given a density matrix ρ, the von Neumann entropy is defined as

$$\boxed{S(\rho) = -\,\mathrm{tr}(\rho \ln \rho) = -\sum_k p_k \ln p_k.} \tag{2.111}$$

Working with the logarithm of an operator can be awkward. That is why in the last equality I expressed $S(\rho)$ in terms of the eigenvalues p_k of ρ. In information theory the last expression in (2.111) is also called the **Shannon entropy** (they usually use the log in base 2, but the idea is the same).

The entropy is seen to be a sum of functions of the form $-p\ln(p)$, where $p \in [0,1]$. The behavior of this function is shown in Fig. 2.3. It tends to zero both when $p \to 0$ and $p \to 1$, and has a maximum at $p = 1/e$. Hence, any state which has $p_k = 0$ *or* $p_k = 1$ will not contribute to the entropy (even though $\ln(0)$ alone diverges, $0\ln(0)$ is well behaved). States that are too deterministic therefore contribute little to the entropy. Entropy likes randomness.

Since each $-p\ln(p)$ is always non-negative, the same must be true for $S(\rho)$:

$$S(\rho) \geq 0. \tag{2.112}$$

Moreover, if the system is in a pure state, $\rho = |\psi\rangle\langle\psi|$, then it will have one eigenvalue $p_1 = 1$ and all others zero. Consequently, in a pure state the entropy will be zero:

$$\boxed{\text{The entropy of a pure state is zero.}} \tag{2.113}$$

In information theory the quantity $-\ln(p_k)$ is sometimes called the *surprise*. When an "event" is rare ($p_k \sim 0$) this quantity is big ("surprise!") and when an event is common ($p_k \sim 1$) this quantity is small ("meh"). The entropy is then interpreted as the *average surprise* of the system, which I think is a little bit funny.

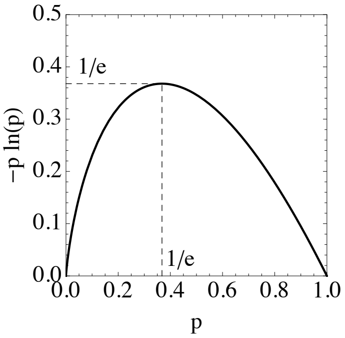

Figure 2.3: The function $-p\ln(p)$, corresponding to each term in the von Neumann entropy (2.111).

As we have just seen, the entropy is bounded from below by 0. But if the Hilbert space dimension d is finite, then the entropy will also be bounded from above. I will leave this proof for you as an exercise. What you need to do is maximize Eq. (2.111) with respect to the p_k, but using Lagrange multipliers to impose the constraint $\sum_k p_k = 1$. Or, if you are not in the mood for Lagrange multipliers, wait until Eq. (2.122) where I will introduce a much easier method to demonstrate the same thing. In any case, the result is

$$\max(S) = \ln(d). \qquad \text{Occurs when} \;\; \rho = \frac{\mathbb{I}}{d}. \tag{2.114}$$

The entropy therefore varies between 0 for pure states and $\ln(d)$ for maximally disordered states. Hence, it clearly serves as a measure of how mixed a state is.

Another very important property of the entropy (2.111) is that it is invariant under unitary transformations:

$$S(U\rho U^\dagger) = S(\rho). \tag{2.115}$$

This is a consequence of the infiltration property of the unitaries $Uf(A)U^\dagger = f(UAU^\dagger)$ [Eq. (1.71)], together with the cyclic property of the trace. Since the time evolution of closed systems are implemented by unitary transformations, this means that the entropy is a constant of motion. We have seen that the same is true for the purity: unitary evolutions do not change the mixedness of a state. Or, in the Bloch sphere picture, unitary evolutions keep the state on the same spherical shell. For open quantum systems this will no longer be the case.

As a quick example, let us write down the formula for the entropy of a qubit. Recall the discussion in Sec. 2.2: the density matrix of a qubit may always be written as in Eq. (2.29). The eigenvalues of ρ are therefore $(1 \pm s)/2$ where $s = \sqrt{s_x^2 + s_y^2 + s_z^2}$ represents the radius of the state in Bloch's sphere. Hence, applying Eq. (2.111) we get

$$S = -\left(\frac{1+s}{2}\right) \ln\left(\frac{1+s}{2}\right) - \left(\frac{1-s}{2}\right) \ln\left(\frac{1-s}{2}\right). \tag{2.116}$$

For a pure state we have $s = 1$ which then gives $S = 0$. On the other hand, for a maximally disordered state we have $s = 0$ which gives the maximum value $S = \ln 2$, the log of the dimension of the Hilbert space. The shape of S is shown in Fig. 2.4.

The quantum relative entropy

Another very important quantity in quantum information theory is the *quantum relative entropy* or *Kullback-Leibler divergence*. Given two density matrices ρ and σ, it is defined as

$$\boxed{S(\rho||\sigma) = \mathrm{tr}(\rho \ln \rho - \rho \ln \sigma).} \tag{2.117}$$

This quantity is important for a series of reasons. But one in particular is that it satisfies the *Klein inequality*:

$$\boxed{S(\rho||\sigma) \geq 0, \qquad S(\rho||\sigma) = 0 \text{ iff } \rho = \sigma.} \tag{2.118}$$

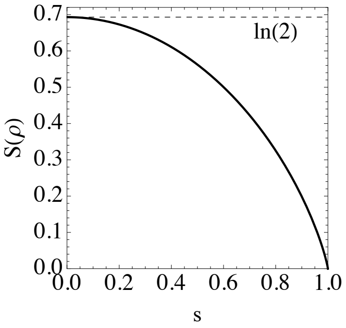

Figure 2.4: The von Neumann entropy for a qubit, Eq. (2.116), as a function of the Bloch-sphere radius s.

The proof of this inequality is really boring and I'm not gonna do it here. You can find it in Nielsen and Chuang or even in Wikipedia.

Eq. (2.118) gives us the idea that we could use the relative entropy as a measure of the *distance* between two density matrices. But that is not entirely precise since the relative entropy does not satisfy the triangle inequality

$$d(x,z) \leq d(x,y) + +d(y,z). \tag{2.119}$$

This is something a true measure of distance must always satisfy. If you are wondering what quantities are actual distances, the *trace distance* is one of them[3]

$$\mathcal{T}(\rho,\sigma) = ||\rho - \sigma||_1 := \text{tr}\left[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right]. \tag{2.120}$$

But there are others as well.

As I mentioned above, the relative entropy is very useful in proving some mathematical relations. For instance consider the result in Eq. (2.114). We can prove it quite easily by noting that

$$\begin{aligned} S(\rho||1/d) &= \text{tr}(\rho\ln\rho) - \text{tr}(\rho\ln(1/d)) \\ &= -S(\rho) + \ln(d). \end{aligned} \tag{2.121}$$

Because of (2.118) we see that

$$S(\rho) \leq \ln(d), \tag{2.122}$$

and $S(\rho) = \ln(d)$ iff $\rho = 1/d$, which is precisely Eq. (2.114). Oh, and by the way, if you felt a bit insecure with the manipulation of $1/d$ in Eq. (2.121), that's ok. The point is that here "1" stands for the identity matrix, but the identity

[3] The fact that $\rho - \sigma$ is Hermitian can be used to simplify this a bit. I just wanted to write it in a more general way, which also holds for non-Hermitian operators.

matrix satisfies the *exact* same properties as the number one, so we can just use the usual algebra of logarithms in this case.

Unlike the entropy, which is always well behaved, the relative entropy may be infinite. The problem is in the last term of (2.117) because we may get a $\ln(0)$ which does not have a 0 in front to save the day. To take an example, suppose ρ is the general state (2.18) and suppose that $\sigma = \text{diag}(f, 1-f)$ for some $f \in [0,1]$. Then

$$\begin{aligned}\text{tr}(\rho \ln \sigma) &= \langle 0|\rho \ln \sigma|0\rangle + \langle 1|\rho \ln \sigma|1\rangle \\ &= p \ln f + (1-p)\ln(1-f).\end{aligned}$$

We can now see that if we happen to have $f = 0$, then the only situation where the first term will not explode is when $p = 0$ as well. This idea can be made mathematically precise as follows. Given a density matrix ρ, we define the **support** of ρ as the vector space spanned by eigenvectors which have non-zero eigenvalues. Moreover, we call the **kernel** as the complementary vector space; that is, the vector space spanned by eigenvectors having eigenvalue zero. Then we can say that $S(\rho||\sigma)$ will be infinite whenever the kernel of σ has an intersection with the support of ρ. If that is not the case, then $S(\rho||\sigma)$ is finite.

Sub-additivity and mutual information

Consider now a bipartite system prepared in a certain state ρ_{AB}. We have seen that if the two systems are not correlated then we can write $\rho_{AB} = \rho_A \otimes \rho_B$. Otherwise, that is not possible. Now we look at the entropy (2.111). When we have two operators separated by a tensor product $\otimes$, the log of the product becomes the sum of the logs:

$$\ln(\rho_A \otimes \rho_B) = (\ln \rho_A) \otimes 1_B + 1_A \otimes (\ln \rho_B). \tag{2.123}$$

This can be viewed more clearly by looking at the eigenvalues of $\rho_A \otimes \rho_B$, which are just of the form $p_k^A p_\ell^B$. Sometimes I'm lazy and I just write this relation as

$$\ln(\rho_A \rho_B) = \ln \rho_A + \ln \rho_B. \tag{2.124}$$

It is then implicit that ρ_A and ρ_B live on separate spaces and therefore commute.

From this it now follows that

$$\begin{aligned}S(\rho_A \otimes \rho_B) &= -\,\text{tr}(\rho_A \rho_B \ln \rho_A) - \text{tr}(\rho_A \rho_B \ln \rho_B) \\ &= -\,\text{tr}_A(\rho_A \ln \rho_A)\,\text{tr}_B(\rho_B) - \text{tr}_B(\rho_B \ln \rho_B)\,\text{tr}_A(\rho_A) \\ &= -\,\text{tr}(\rho_A \ln \rho_A) - \text{tr}(\rho_B \ln \rho_B).\end{aligned}$$

I know this calculation is a bit messy, but please try to convince yourself that it's ok. For instance, you can do everything with $\otimes$ and use Eq. (2.65). In any case, what we conclude is that

$$S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B). \tag{2.125}$$

Thus, the entropy is an *additive* quantity: if two systems are uncorrelated, the total entropy is simply the sum of the parts.

This is no longer true if ρ_{AB} is a correlated state. In fact, the entropy of ρ_{AB} is related to the entropy of the reduced density matrices $\rho_A = \text{tr}_B\, \rho_{AB}$ and $\rho_B = \text{tr}_A\, \rho_{AB}$ by the **subadditivity condition**

$$\boxed{S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).} \tag{2.126}$$

where the equality holds only for a product state $\rho_{AB} = \rho_A \otimes \rho_B$. Another way to write this is as $S(\rho_{AB}) \leq S(\rho_A \otimes \rho_B)$. This has a clear interpretation: by taking the partial trace we *loose information* so that the entropy afterwards is larger.

The proof of Eq. (2.126) can be done easily using the relative entropy. We just need to convince ourselves that

$$S(\rho_{AB}||\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \tag{2.127}$$

Then, because of (2.118), this quantity will always be non-negative. So let's do it: let's see that (2.127) is indeed correct.

$$\begin{aligned} S(\rho_{AB}||\rho_A \otimes \rho_B) &= = \text{tr}(\rho_{AB} \ln \rho_{AB}) - \text{tr}(\rho_{AB} \ln \rho_A \rho_B) \\ &= -S(\rho_{AB}) - \text{tr}(\rho_{AB} \ln \rho_A) - \text{tr}(\rho_{AB} \ln \rho_B). \end{aligned} \tag{2.128}$$

Now comes the key point: given any operator $\mathcal{O}$ we can always take the trace in steps:

$$\text{tr}(\mathcal{O}) = \text{tr}_A(\text{tr}_B(\mathcal{O})).$$

Then, to deal with $\text{tr}(\rho_{AB} \ln \rho_A)$ we can first take the trace in B. This will only affect ρ_{AB} and it will turn it into ρ_A:

$$\text{tr}(\rho_{AB} \ln \rho_A) = \text{tr}_A\, \rho_A \ln \rho_A.$$

This is always true, even when ρ_{AB} is not a product. Plugging this in (2.128), we immediately see that (2.127) will hold.

Looking back now at Eqs. (2.126) and (2.127) we see that we have just found a quantity which is always non-negative and is zero exactly when the two systems are uncorrelated ($\rho_{AB} = \rho_A \otimes \rho_B$). Thus, we may use this as a quantifier of the **total degree of correlations**. We call this quantity the **mutual information**:

$$\boxed{I(\rho_{AB}) := S(\rho_{AB}||\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \geq 0.} \tag{2.129}$$

This is one the central concepts in all of quantum information. *It represents the amount of information stored in AB which is not stored in A and B, when*

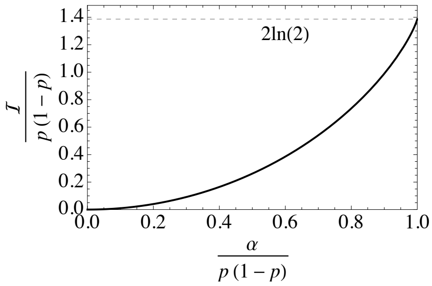

Figure 2.5: Eq. (2.131) plotted in terms of convenient quantities.

taken separately. One thing I should warn, though, is that the mutual information quantifies the *total* degree of correlations, in the sense that it does not distinguish between classical and quantum contributions. A big question in the field is how to separate the mutual information in a quantum and a classical part. We will get back to that later.

Let me try to give an example of the mutual information. This is always a little bit tricky because even for two qubits, the formulas can get quite ugly (although asking Mathematica to write them down is really easy). So for the purpose of example, let us consider a state of the form:

$$\rho_{AB} = \begin{pmatrix} p^2 & 0 & 0 & 0 \\ 0 & p(1-p) & \alpha & 0 \\ 0 & \alpha & p(1-p) & 0 \\ 0 & 0 & 0 & (1-p)^2 \end{pmatrix}. \tag{2.130}$$

This has the structure of Eq. (2.79), but with ρ_A and ρ_B being equal and diagonal: $\rho_A = \rho_B = \text{diag}(p, 1-p)$. The constant α here is also not arbitrary, but is bounded by $|\alpha| < p(1-p)$, which is a condition so that the eigenvalues of ρ_{AB} are always non-negative. The mutual information is then

$$\mathcal{I}(\rho_{AB}) = p(1-p)\ln\left[\frac{p^2(1-p)^2-\alpha^2}{p^2(1-p^2)}\right] + \alpha\ln\left[\frac{p(1-p)+\alpha}{p(1-p)-\alpha}\right]. \tag{2.131}$$

This function is plotted in Fig. 2.5. As expected, the larger the correlation α, the larger is the mutual information. The maximum value occurs when $|\alpha| = p(1-p)$ and has the value $\mathcal{I} = 2p(1-p)\ln(2)$.

Next suppose that $\rho_{AB} = |\psi\rangle\langle\psi|$ is actually a pure state. Then $S(\rho_{AB}) = 0$. Moreover, we have seen in Sec. 2.8 that the reduced density matrices of A and B can both be written in diagonal form in terms of the Schmidt coefficients, Eqs. (2.100) and (2.101). Thus, it follows that in this case

$$S(\rho_A) = S(\rho_B) \qquad \text{when } \rho_{AB} \text{ is pure.} \tag{2.132}$$

Hence, the mutual information becomes

$$\mathcal{I}(\rho_{AB}) = 2S(\rho_A) = 2S(\rho_B) \qquad \text{when } \rho_{AB} \text{ is pure.} \tag{2.133}$$

We therefore conclude that for pure states the maximum amount of information stored in non-local correlations is twice the information of each of the parts.

For the case of pure states, we saw that we could quantify the degree of entanglement by means of the purity of ρ_A or ρ_B. Another way to quantify entanglement is by means of the entropy $S(\rho_A)$ and $S(\rho_B)$. For this reason, this is sometimes referred to as the **entanglement entropy**. Eq. (2.133) then shows us that for pure states the mutual information is twice the entanglement entropy. On the other hand, if the state is not pure, than entanglement will be mixed with classical correlations. An important question is then what part of $\mathcal{I}$ is due to entanglement and what part is classical. We will get back to this later in the course.

In addition to the subadditivity inequality (2.126), the von Neumann entropy also satisfies the **strong subadditivitiy inequality**:

$$\boxed{S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).} \tag{2.134}$$

If B is a Hilbert space of dimension 1 this reduces to Eq. (2.126). The intuition behind this formula is as follows (Preskill): We can think as AB and BC as two overlapping systems, so that $S(\rho_{ABC})$ is the entropy of their union and $S(\rho_B)$ is the entropy of their intersection. Then Eq. (2.134) says this cannot exceed the sum of the entropies of the parts. Eq. (2.134) can also be stated in another way as

$$S(\rho_A) + S(\rho_B) \leq S(\rho_{AC}) + S(\rho_{BC}). \tag{2.135}$$

The strong subadditivity inequality turns out to be an essential property in quantum information tasks, such as communication protocols. The proof of Eq. (2.134), however, turns out to be quite difficult and we will not be shown here. You can find it, for instance, in Nielsen and Chuang, chapter 11.

Convexity of the entropy

Consider now a bipartite state of the form

$$\rho_{AB} = \sum_i p_i \rho_i \otimes |i\rangle\langle i|, \tag{2.136}$$

where ρ_i are valid density matrices and p_i are arbitrary probabilities. This type of state is what we call a *quantum-classical state*. It is like a mixture of classical probabilities from the point of view of B, but with (possibly) quantum density matrices from the point of view of A. That can be seen more clearly by looking at the reduced density matrices:

$$\rho_A = \operatorname{tr}_B \rho_{AB} = \sum_i p_i \rho_i, \tag{2.137}$$

$$\rho_B = \operatorname{tr}_A \rho_{AB} = \sum_i p_i |i\rangle\langle i|. \tag{2.138}$$

Each ρ_i may have quantum stuff inside them and what we are doing in ρ_A is making classical mixtures of these guys.

The entropy of ρ_B is now the **classical Shannon entropy** of the probability distribution p_i:

$$S(\rho_B) := H(p_i) = -\sum_i p_i \ln p_i. \tag{2.139}$$

The use of the letter H is not completely necessary. I just put it there to emphasize that we are talking about the entropy of a set of numbers $\{p_i\}$ and not a density matrix. Next let us compute the entropy of ρ_{AB}. Denote by $P_{i,j}$ the j-th eigenvalue of each ρ_i. Then the eigenvalues of ρ_{AB} will be $p_i P_{i,j}$. Thus

$$S(\rho_{AB}) = -\sum_{i,j} p_i P_{i,j} \ln(p_i P_{i,j}) \tag{2.140}$$

$$= -\sum_{i,j} p_i P_{i,j} \ln p_i - \sum_{i,j} p_i P_{i,j} \ln P_{i,j} \tag{2.141}$$

In the first term we now use $\sum_j P_{i,j} = 1$, which is the normalization condition for each ρ_i. What is left is then $S(\rho_B) = S(p_i)$. In the second term, on the other hand, we note that for each i, the sum over j is just $S(\rho_i) = -\sum_j P_{i,j} \ln P_{i,j}$. Thus we finally get

$$S(\rho_{AB}) = H(p_i) + \sum_i p_i S(\rho_i). \tag{2.142}$$

We therefore see that in this case the total entropy has two clear contributions. The first is the disorder introduced by the probability distribution p_i and the second is the local disorder contained in each ρ_i.

Using now the subadditivity formula (2.126) together with the fact that $S(\rho_B) = H(p_i)$, we also see that

$$S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i), \tag{2.143}$$

where I used the form of ρ_A in Eq. (2.137). The entropy is therefore a concave function of its arguments. The logic behind this formula is that $\sum_i p_i \rho_i$ contains not only ignorance about the ρ_i but also about the p_i. So its total entropy must be higher than the sum of the parts.

Eq. (2.143) provides a lower bound to the entropy of mixtures. It turns out that it is also possible to find an upper bound, so that instead of (2.143) we can write the more general result

$$\boxed{H(p_i) + \sum_i p_i S(\rho_i) \geq S\Big(\sum_i p_i \rho_i\Big) \geq \sum_i p_i S(\rho_i).} \tag{2.144}$$

The proof is given in chapter 11 of Nielsen and Chuang. The cool thing about this new bound is that it allows for an interpretation of entropy in terms of the **ambiguity of mixtures**. Remember that we discussed how the same density matrix ρ could be constructed from an infinite number of combinations of pure states

$$\rho = \sum_i q_i |\psi_i\rangle\langle\psi_i|. \tag{2.145}$$

In this formula there can be an arbitrary number of terms and the $|\psi_i\rangle$ do not need to be orthogonal or anything. All we require is that the q_i behave like probabilities. Hence, due to this flexibility, there is an infinite number of choices for $\{q_i, |\psi_i\rangle\}$ which give the same ρ. But note how this falls precisely into the category of Eq. (2.144), with $\rho_i = |\psi_i\rangle\langle\psi_i|$ and $p_i \to q_i$. Since $S(\rho_i) = 0$ for a pure state, we then find that

$$\boxed{S(\rho) \leq H(q_i).} \tag{2.146}$$

That is, *the von Neumann entropy is the entropy that minimizes the classical distribution of the probabilities* $H(q_i)$. In terms of the eigenvalues p_k of ρ, we have $S(\rho) = -\sum_k p_k \ln p_k$ so that the equality in Eq. (2.146) is obtained when the mixture is precisely that of the eigenvalues/eigenvectors of ρ.

Rényi entropy

A generalization of the von Neumann entropy that is also popular in quantum information are the so-called Rényi entropies, defined as

$$\boxed{S_\alpha(\rho) = \frac{1}{1-\alpha} \ln \operatorname{tr} \rho^\alpha.} \tag{2.147}$$

where α is a tunable parameter in the range $[0, \infty)$. This therefore corresponds to a continuous family of entropies. I particularly like $\alpha = 2$, which is simply minus the logarithm of the purity:

$$S_2(\rho) = -\ln \operatorname{tr} \rho^2. \tag{2.148}$$

Another special case is $\alpha = 1$, where we recover the von Neumann entropy. Note how this is tricky because of the denominator in Eq. (2.147). The safest way

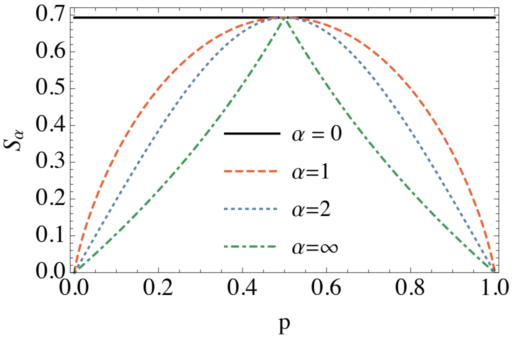

Figure 2.6: The Rényi entropies for a 2-state system, computed using Eq. (2.152) for different values of α.

to do this is to expand x^α in a Taylor series in α around $\alpha = 1$. We have the following result from introductory calculus:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} x^\alpha = x^\alpha \ln(x).$$

Thus, expanding x^α around $\alpha = 1$ we get:

$$x^\alpha \simeq x^1 + x^1 \ln(x)(\alpha - 1).$$

Now we substitute this into Eq. (2.147) to get

$$\begin{aligned} S_\alpha(\rho) &\simeq \frac{1}{1-\alpha} \ln\left\{ \operatorname{tr}\rho + (\alpha - 1)\operatorname{tr}(\rho \ln \rho) \right\} \\ &= \frac{1}{1-\alpha} \ln\left\{ 1 + (\alpha - 1)\operatorname{tr}(\rho \ln \rho) \right\}. \end{aligned}$$

Since we want the limit $\alpha \to 1$, we may expand the logarithm above using the formula $\ln(1 + x) \simeq x$. The terms $\alpha - 1$ will then cancel out, leaving us with

$$\lim_{\alpha \to 1} S_\alpha(\rho) = -\operatorname{tr}(\rho \ln \rho), \tag{2.149}$$

which is the von Neumann entropy. The Rényi entropy therefore forms a family of entropies which contains the von Neumann entropy as a particular case. Other particular cases of importance are $\alpha = 0$, which is called the *max entropy*, and $\alpha = \infty$ which is called the *min entropy*. Using the definition (2.147) we see that

$$S_0(\rho) = \ln(d), \tag{2.150}$$

$$S_\infty(\rho) = -\ln \max_k p_k. \tag{2.151}$$

As an example, consider a qubit with eigenvalues p and $1-p$. Then $\mathrm{tr}(\rho^\alpha) = p^\alpha + (1-p)^\alpha$ so that Eq. (2.147) becomes

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \ln \Big\{ p^\alpha + (1-p)^\alpha \Big\}. \tag{2.152}$$

This result is plotted in Fig. 2.6 for several values of α. As can be seen, except for $\alpha \to 0$, which is kind of silly, the behavior of all curves is qualitatively similar.

Integral representations of $\ln(\rho)$

When dealing with more advanced calculations, sometimes dealing with $\ln(\rho)$ in terms of eigenvalues can be hard. An alternative is to write the logarithm of operators as an integral representation. I know two of them. If you know more, tell me and I can add them here. A simple one is

$$\ln(\rho) = (\rho - 1) \int\limits_0^1 \frac{\mathrm{d}x}{1 + x(\rho - 1)}. \tag{2.153}$$

Here whenever ρ appears in the denominator, what is meant is the matrix inverse. Another formula is[4]

$$\ln(\rho) = (\rho - 1) \int\limits_0^\infty \frac{\mathrm{d}x}{(1+x)(\rho+x)} \tag{2.154}$$

$$= \int\limits_0^\infty \mathrm{d}x \left(\frac{1}{1+x} - \frac{1}{\rho+x} \right). \tag{2.155}$$

This last formula in particular can now be used as the starting point for a series expansion, based on the matrix identity

$$\frac{1}{A+B} = \frac{1}{A} - \frac{1}{A} B \frac{1}{A+B}. \tag{2.156}$$

For instance, after one iteration of Eq. (2.155) we get

$$\ln(\rho) = \int\limits_0^\infty \mathrm{d}x \left(\frac{1}{1+x} - \frac{1}{x} + \frac{\rho}{x^2} - \frac{\rho^2}{x^2} \frac{1}{\rho + x} \right). \tag{2.157}$$

2.10 Generalized measurements and POVMs

So far our discussion of measurements has been rather shallow. What I have done so far is simply postulate the idea of a projective measurement, without

[4] See M. Suzuki, *Prog. Theo. Phys*, **100**, 475 (1998)