Lecture 4 – Entropies

Peter Brown

October 16, 2022

Last week we saw

- States: Density matrices $\rho \in \mathcal{D}(\mathcal{H})$.
- Transformations: CPTP maps $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$
- Measurements: Projective / POVMs / instruments

Together with the various characterizations of these objects. For example using dilation theorems we could always consider our system as part of a larger system and recover the "closed system" formalism of quantum theory.

Before we move on let us note some notation. The logarithm log will be taken base 2. If we have two random variables X,Y over finite sets $\mathcal{X}$, $\mathcal{Y}$, we denote the joint probability distribution of X and Y by p_{XY}, i.e., $p_{XY}(x,y)$ is the probability that $X = x$ and $Y = y$. We denote the *marginal* distributions by $p_X(x) = \sum_y p_{XY}(x,y)$ and $p_Y(y) = \sum_x p_{XY}(x,y)$. Furthermore we denote the conditional distribution of random variable X given random variable Y by $p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$.

Note that we will always assume that random variables take values from finite sets and similarly we always assume that the dimension of Hilbert spaces are finite.

1 Classical entropies

Entropies, in general, represent a notion of uncertainty about some random process. Given a random variable X which takes values in some set $\mathcal{X}$ with probability p_X we can define the *surprisal* function $S : \mathcal{X} \to \mathbb{R}$ as

$$S(x) := -\log p_X(x)\,. \tag{1}$$

The surprisal attempts to quantify how surprised we are when we see the outcome x of our random variable X. If the outcome x has a high probability then the surprisal is small, whereas if the outcome x has a low probability then the we are very surprised when we see it happen and indeed $S(x) \to \infty$ as $p(x) \to 0$ (see Figure 1). Note that the surprisal is also additive for independent random variables. So if we observe two independent random events then our total surprisal is the sum of the individual surprisals. In certain contexts one can also view the surprisal as the amount of information gained upon learning the outcome x.

The *Shannon entropy* of the random variable X denoted $H(X)$ is then the expected surprisal

$$H(X) := -\sum_x p(x) \log p(x) = \mathbb{E}[S(X)], \tag{2}$$

in other words, on average how surprised are we at the outcome of X. Note that if an outcome has 0 probability of occurring then we take the convention $0 \log 0 = 0$.

Example 1.

1. If we have a deterministic distribution, i.e., there exists an $x^* \in \mathcal{X}$ such that $p(x^*) = 1$ then we have

$$H(X) = -\sum_{x:p(x)>0} p(x) \log p(x) = -1 \log 1 = 0$$

you're not very surprised if your random variable is deterministic!

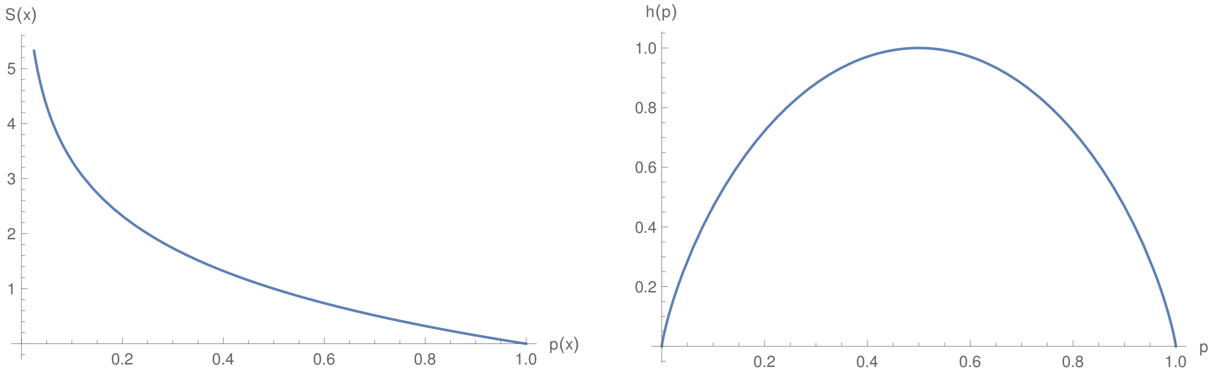

Figure 1: In the first plot we see the surprisal as a function of the probability $p(x)$. Note $S(x) = 0$ when $p(x) = 1$ and $S(x) \to \infty$ as $p(x) \to 0^+$. In the second plot we see the binary entropy as a function of $p(0)$.

2. Suppose you have a binary random variable X, so $\mathcal{X} = \{0, 1\}$ then the distribution is specified fully by the probability $p(0) := p \in [0, 1]$. In that case we have $H(X) := h(p)$ where

$$h(p) := -p \log p - (1 - p) \log(1 - p) \, .$$

in Figure 1 we see a plot of the binary entropy $h(p)$.

We can also define a conditional entropy. If we have two random variables X and Y that have some joint distribution p_{XY} then the *conditional Shannon entropy* $H(X|Y)$ is defined as

$$H(X|Y) := H(XY) - H(Y) \, , \tag{3}$$

where $H(XY) = -\sum_{xy} p_{XY}(x, y) \log p_{XY}(x, y)$. Again one can roughly interpret this as the average uncertainty one has about X given that one access to Y. Finally, we can define the *mutual information* between two random variables X and Y as

$$I(X : Y) = H(X) + H(Y) - H(X, Y) \, , \tag{4}$$

roughly the mutual information describes how dependent two random variables are. Again these rough interpretations can be made more concrete when we find these functions as characterizing how well we can perform some information theoretic task like compression or communication. We will explore this more in a later lecture but for now we will just treat these functions with their rough interpretations and understand how to compute them and the various properties that they possess.

Lemma 1 (Properties of classical entropies)**.** *Let X, Y and Z be random variables. Then*

1. *If X and Y are independent then*

   (a)
   $$H(XY) = H(X) + H(Y) \tag{5}$$

   (b)
   $$H(X|Y) = H(X) \tag{6}$$

   (c)
   $$I(X : Y) = 0 \tag{7}$$

2. *For all X, Y we have*
   $$H(X|Y) = \sum_y p_Y(y) H(X|Y = y) \tag{8}$$
   where $H(X|Y = y) = -\sum_x p(x|y) \log p(x|y)$.

3. *Chain rules:*
   $$H(X, Y) = H(X) + H(Y|X) \tag{9}$$
   and its generalization
   $$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \, . \tag{10}$$

4. *Extra information doesn't increase uncertainty*

$$I(X:Y) \geq 0 \tag{11}$$

and as a consequence

$$H(X|Y) \leq H(X)\,. \tag{12}$$

5. *Minima / maxima: we have*

$$\begin{aligned} 0 \leq H(X) &\leq \log(|\mathcal{X}|) \\ 0 \leq H(X|Y) &\leq \log(|\mathcal{X}|) \\ 0 \leq I(X:Y) &\leq \min\{H(X), H(Y)\} \end{aligned} \tag{13}$$

the first inequality is achieved for a uniform distribution on X, the second is achieved whenever X is uniform given Y and the latter occurs whenever $H(X|Y) = 0$ or $H(Y|X) = 0$.

1.1 The min-entropy

There are other notions of entropy beyond the Shannon entropy that capture other notions of uncertainty. One such quantity is called the min-entropy (denoted $H_{\min}$). This quantity is particularly suited to cryptography and roughly captures how many random bits can be extracted from some source. For instance consider the following example.

Example 2 (Guessing probability).
Suppose you have a random variable X that produces n-bit strings $x_1x_2\ldots x_n$ with a distribution

$$p_X(x_1x_2\ldots x_n) = \begin{cases} 1/2 & \text{if } x_1x_2\ldots x_n = 00\ldots0 \\ \frac{1}{2(2^n-1)} & \text{otherwise} \end{cases}\,. \tag{14}$$

If you wanted to use this random variable as a source of randomness, to say seed a secret key, you would want that the probability that you can guess the outcome of the random variable is very small (otherwise someone could guess your seed and your protocol is no longer secure). In this case the probability that some adversary can guess the outcome of your random variable X is given by

$$P_{\text{guess}}(X) = \max_x p_X(x) = \frac{1}{2} \tag{15}$$

because her best strategy is to just always guess that the outcome will be the all 0 bitstring and she will guess correctly 1/2 of the time. In this context, we can think of this random variable X as being no more useful that a single uniformly[1] random bit Y, $p_Y(0) = p_Y(1) = 1/2$.

If we look at the Shannon entropy of the distribution then we find that

$$H(X) = 1 + \frac{1}{2}\log(2^n - 1) \tag{16}$$

that is the Shannon entropy does not capture the very pessimistic viewpoint of cryptography and would say that the randomness grows linearly with the length of the bitstring n. It is precisely the Shannon entropy's definition as an average uncertainty which means it fails to capture this worst-case uncertainty that we are looking for in this scenario.[2]

In light of the above example we introduce a new entropy called the *min-entropy* which is defined for a random variable X with distribution p_X as

$$H_{\min}(X) := -\log(\max_x p_X(x))\,. \tag{17}$$

Note the quantity inside the logarithm is the maximum probability with which someone (without additional information) can guess the outcome of your random variable X (denoted $P_{\text{guess}}(X)$). The min-entropy $H_{\min}(X)$ roughly captures the number of uniformly random bits that one can transform the output of the source X into.[3]

[1] Recall that a distribution p_X is *uniform* if all the outcomes are equally likely, i.e. $p_X(x) = \frac{1}{|\mathcal{X}|}$.

[2] There is no 'one entropy to rule them all'. In fact there are numerous different entropies that are all suited to different tasks. In this case we need a new entropy that captures a worst-case uncertainty.

[3] This procedure is called *randomness extraction* and is used extensively in quantum cryptography.

We can also define a *conditional min-entropy* of X given another random variable Y as

$$H_{\min}(X|Y) = -\log(P_{\text{guess}}(X|Y)) \tag{18}$$

where

$$P_{\text{guess}}(X|Y) := \sum_y p_Y(y) \max_x p_{X|Y}(x|y) \,. \tag{19}$$

Again $P_{\text{guess}}(X|Y)$ should be interpreted as a guessing probability, although now it is the maximum probability with which someone who has access to the random variable Y can guess the outcome of the random variable X. Indeed there best strategy is to sample Y receive and outcome y and then proceed to guess the most likely outcome of X given they received y which is $\text{argmax}_x p_{X|Y}(x|y)$. Finally one can interpret again $H_{\min}(X|Y)$ as the number of uniformly random bits that one can transform the output of X into, except now these bits will remain uniformly distributed even from the perspective of someone who has access to Y. The conditional min-entropy is particularly useful in cryptography when we want to assume that an adversary may have additional information about X in the form of some random variable Y.

2 Quantum entropies

We now move onto define quantum entropies. These new entropies will generalize the notion of the classical entropies that we have already seen and due to the richer nature of density matrices we will see new properties of these entropies that don't appear in the classical case. In a later lecture we will also see some information theoretic tasks which give these entropies their operational interpretations. We begin with the quantum generalization of the Shannon entropy.[4]

Definition 1 (von Neumann entropy)**.** Let A be a quantum system and let $\rho \in \mathcal{D}(A)$ then the *von Neumann entropy* of the system A in the state ρ is given by

$$H(A)_\rho := -\text{Tr}\left[\rho \log \rho\right] \,. \tag{20}$$

Remark 1.

1. Recall from the previous week's notes we can use the spectral theorem to define matrix functions. In this case if we write the spectral decomposition of $\rho_A \in \mathcal{D}(A)$,

$$\rho_A = \sum_i \lambda_i |v_i\rangle\langle v_i|$$

   then

$$\rho_A \log \rho_A = \sum_i \lambda_i \log(\lambda_i) |v_i\rangle\langle v_i| \,.$$

   Therefore,

$$H(A)_\rho = -\sum_i \lambda_i \log(\lambda_i) \tag{21}$$

   where like in the case of the Shannon entropy we treat $0 \log 0 = 0$. In particular the von Neumann entropy is actually just the Shannon entropy of the eigenvalues of ρ (which form a probability distribution as ρ is a density matrix).

2. In the definition we subscript $H(A)$ with the state ρ which we use to calculate the entropy. However, if ρ is clear from context or not explicitly needed then we will be lazy and drop the subscript. In principle this subscripting could also be done for the classical entropies as well.

Example 3.

1. Suppose we have a pure state $\rho_A = |\psi\rangle\langle\psi|$. Then

$$H(A)_\rho = \text{Tr}\left[1 \log(1) |\psi\rangle\langle\psi|\right] = 0 \,. \tag{22}$$

   The entropy is always 0 for pure states (cf. deterministic distributions in the classical case).

[4] Curiously the von Neumann entropy was actually defined (1927) before the Shannon entropy (1948) and the story goes that von Neumann told Shannon to call his quantity an entropy because no-one really knows what entropy is and so if you're in a debate with someone then you'll always have the upper hand.

2. Consider a qubit system A in the state $\rho = p|0\rangle\langle 0| + (1-p)|+\rangle\langle +|$. As a matrix in the computational basis we have
$$\rho = \begin{pmatrix} \frac{1+p}{2} & \frac{1-p}{2} \\ \frac{1-p}{2} & \frac{1-p}{2} \end{pmatrix} . \tag{23}$$
It has a characteristic polynomial
$$\det(\rho - \lambda I) = \left(\frac{1+p}{2} - \lambda\right)\left(\frac{1-p}{2} - \lambda\right) - \frac{(1-p)^2}{4} = \lambda^2 - \lambda + \frac{p(1-p)}{2} .$$
therefore we have eigenvalues $\lambda = \frac{1 \pm \sqrt{1-2p(1-p)}}{2}$ and then
$$H(A)_\rho = h\left(\frac{1+\sqrt{1-2p(1-p)}}{2}\right) \tag{24}$$
where h is the binary entropy. For $p = 0$ or $p = 1$ the state is pure and so the entropy vanishes. The largest entropy occurs in this case at $p = 1/2$ in which case we have $H(A) \approx 0.6$.

Definition 2. Let A and B be quantum systems and let $\rho_{AB} \in \mathcal{D}(AB)$. We define the *conditional von Neumann entropy* of system A given system B for the state ρ_{AB} as
$$H(A|B)_{\rho_{AB}} := H(AB)_{\rho_{AB}} - H(B)_{\rho_B} \tag{25}$$
where as usually $\rho_B = \mathrm{Tr}_A\left[\rho_{AB}\right]$.

Whilst in the case of classical entropies we always had $H(X|Y) \geq 0$ it is possible now that $H(A|B)_{\rho_{AB}} < 0$ (see for example Exercise 13). This property of negative conditional entropy is intricately linked to entanglement.

We can also define the *quantum mutual information* in a similar way.

Definition 3 (Quantum mutual information)**.** Let A and B be quantum systems and let $\rho_{AB} \in \mathcal{D}(AB)$. Then the *mutual information* between systems A and B when in the state ρ_{AB} is defined as
$$I(A:B)_{\rho_{AB}} := H(A)_{\rho_A} + H(B)_{\rho_B} - H(AB)_{\rho_{AB}} . \tag{26}$$

Lemma 2 (Basic properties of quantum entropies)**.** *Let A, B and C be quantum systems and let $\rho_{ABC} \in \mathcal{D}(ABC)$.*

1. *If $\rho_{AB} = \rho_A \otimes \rho_B$ (systems are independent) then*

   (a)
$$H(AB) = H(A) + H(B) \tag{27}$$
   (b)
$$H(A|B) = H(A) \tag{28}$$
   (c)
$$I(A:B) = 0 \tag{29}$$

2. *If the system B is* classical*, i.e., $\rho_{AB} = \sum_b p_B(b)\rho_A(b) \otimes |b\rangle\langle b|$, then*
$$H(A|B) = \sum_b p_B(b) H(A|B=b)_{\rho_A(b)} . \tag{30}$$

3. *Chain rules*
$$H(A,B) = H(A) + H(B|A) \tag{31}$$
   and its generalization
$$H(AB|C) = H(A|C) + H(B|AC) . \tag{32}$$

4. *Extra information doesn't increase uncertainty*
$$I(A:B) \geq 0 \tag{33}$$
and as a consequence
$$H(A|B) \leq H(A). \tag{34}$$

5. *Minima/maxima: we have*
$$\begin{aligned} 0 \leq H(A) &\leq \log(d_A) \\ -\log(d_A) \leq H(A|B) &\leq \log(d_A) \\ 0 \leq I(A:B) &\leq 2\log\min\{d_A, d_B\} \end{aligned} \tag{35}$$
where d_A and d_B are the dimension of system A and B respectfully.

Proof. We prove 1, 2 and 3 but 4 and 5 require extra results beyond the scope of this course so we will ignore them.

1. Let the eigenvalues of ρ_A be $\{\mu_i\}_i$ and let the eigenvalues of ρ_B be $\{\nu_j\}_j$ then the eigenvalues of $\rho_A \otimes \rho_B$ are $\{\mu_i\nu_j\}_{i,j}$. Then
$$H(AB) = -\sum_{ij}\mu_i\nu_j\log(\mu_i\nu_j) = -\sum_{ij}\mu_i\nu_j(\log(\mu_i)+\log(\nu_j)) = -\sum_i \mu_i\log(\mu_i) - \sum_j \nu_j\log(\nu_j) = H(A)+H(B).$$
where on the 3rd equality we used the fact that $\sum_i \mu_i = \sum_j \nu_j = 1$.

   Parts (b) and (c) follow immediately from part (a).

2. We have the qc-state $\rho_{AB} = \sum_b p_B(b)\rho_A(b) \otimes |b\rangle\langle b|$. If $\rho_A(b) = \sum_i \lambda_{i,b}|v_{i,b}\rangle\langle v_{i,b}|$ is the spectral decomposition of $\rho_A(b)$ then the spectral decomposition of ρ_{AB} is
$$\rho_{AB} = \sum_b\sum_i p_B(b)\lambda_{i,b}|v_{i,b}\rangle\langle v_{i,b}| \otimes |b\rangle\langle b|.$$
Hence it's eigenvalues are $\{p_B(b)\lambda_{i,b}\}_{i,b}$. Therefore
$$\begin{aligned} H(AB) &= -\sum_{i,b} p_B(b)\lambda_{i,b}\log(p_B(b)\lambda_{i,b}) \\ &= -\sum_b p_B(b)\log p_B(b) - \sum_b p_B(b)\sum_i \lambda_{i,b}\log\lambda_{i,b} \\ &= H(B)_{\rho_B} + \sum_b p_B(b)H(A|B=b)_{\rho_A(b)}\,. \end{aligned}$$
The result then follows from $H(A|B) = H(AB) - H(B)$.

3. This proof is the same as the classical case. The first follows immediately from the definition of conditional entropy $H(B|A) = H(AB) - H(A)$. The second can be derived as
$$H(AB|C) = H(ABC) - H(C) = H(ABC) + H(AC) - H(AC) - H(C) = H(A|C) + H(B|AC)\,. \tag{36}$$

□

2.1 Quantum min-entropy

Just like in classical case the entropies above have an ‘average case’ flavour to them. We can again define a worst case entropy called the min-entropy. However, in order to keep things simple we will only consider the case of cq-states – for future topics we will discuss these are the relevant quantities anyway

Definition 4. Consider the cq-state $\rho_{XB} = \sum_x p_X(x)|x\rangle\langle x| \otimes \rho_B(x)$ where X is a classical system. Then we define the *conditional min-entropy* of X given B as
$$H_{\min}(X|B) := -\log P_{\text{guess}}(X|B) \tag{37}$$
where
$$P_{\text{guess}}(X|B) := \max_{\text{POVMs }\{M_x\}_x}\sum_x p_X(x)\text{Tr}\left[M_x\rho_B(x)\right] \tag{38}$$

The function $P_{\text{guess}}(X|B)$ can again be interpreted as a guessing probability. However now, instead of have access to a random variable Y, we have access to a quantum system B. The strategy to guess the random variable X will be to perform some measurement $\{M_x\}_x$ on our quantum system B and then when we receive the outcome x we will guess that $X = x$. The probability that we succeed using this strategy is then

$$\begin{aligned}\sum_x \mathbb{P}[X = x,\ B \text{ measures } x] &= \sum_x \mathbb{P}[X = x]\mathbb{P}[B \text{ measures } x \,|\, X = x] \\ &= \sum_x p_X(x)\mathrm{Tr}\left[\rho_B(x) M_x\right] .\end{aligned}$$

If we then maximize over all such measurements we will find the best guessing probability.

3 Exercises

Note: For questions that require the computation of eigenvalues or eigenvectors for matrices larger than a qubit please feel free to use a computer rather than doing it by hand.

1. Consider the probability distribution
$$p_X(x) = \begin{cases} 1/4 & \text{if } x = 0 \\ 1/8 & \text{if } x = 1 \\ 5/8 & \text{if } x = 2 \end{cases} \tag{39}$$
Compute $H(X)$ and $H_{\min}(X)$.

2. Consider to binary random variables X and Y which have a joint distribution
$$p_{XY}(x,y) = \begin{cases} \frac{1-p}{2} & \text{if } (x,y) \in \{(0,0),(1,1)\} \\ \frac{p}{2} & \text{if } (x,y) \in \{(0,1),(1,0)\} \end{cases} \tag{40}$$
where $p \in [0, 1/2]$
   (a) Compute $H(X|Y)$.
   (b) Compute $H_{\min}(X|Y)$.
   (c) Compute $I(X : Y)$.

3. Let X and Y be random variables. Prove that $H(X|Y) \geq H_{\min}(X|Y)$.

4. Consider a qubit system A in a state $\rho_A = \begin{pmatrix} 3/4 & 1/8 \\ 1/8 & 1/4 \end{pmatrix}$, compute $H(A)_\rho$.

5. Consider a two-qubit system AB in a state
$$\rho_{AB} = p|\Phi^+\rangle\langle\Phi^+| + (1-p)|\Psi^+\rangle\langle\Psi^+| .$$
where $|\Phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ and $|\Psi^+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$.
   (a) Compute $H(A|B)$. (Note: computing the spectral decomposition of ρ_{AB} may be easier than you think).
   (b) Compute $I(A : B)$.

6. For each of the inequalities in Equation (35) find a quantum state that saturates the inequality (i.e., achieves equality).

7. Let $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$ be a pure bipartite quantum state. Show that
$$H(A) = H(B) . \tag{41}$$

8. Let $\rho_{ABC} = |\psi\rangle\langle\psi|_{ABC}$ be a pure tripartite state. Show that
$$H(A|B) = -H(A|C) \tag{42}$$

9. Consider a cq-state $\rho_{XB} = \sum_x p_X(x)|x\rangle\langle x| \otimes \rho_B(x)$. Show that

$$H(X|B) = H(X)_{p_X} + \sum_x p_X(x) H(B|X = x)_{\rho_B(x)} - H(B)_{\rho_B} . \tag{43}$$

10. Let Q_A and Q_B be qubit systems with Alice holding Q_A and Bob holding Q_B. Furthermore let the state of the two-qubit system be $\rho_{Q_A Q_B} \in \mathcal{D}(Q_A Q_B)$. Suppose Alice performs a measurement $\{M_a\}_a$ on her part of the system and she records the outcome of that measurement in some classical system A. After the measurement we can consider the classical-quantum system AQ_B which represents the joint state of Alice's measurement outcome and Bob's quantum system,

$$\rho_{AQ_B} = \sum_a p_A(a)|a\rangle\langle a| \otimes \rho_{Q_B}(a) . \tag{44}$$

where $p_A(a)$ is the probability that Alice receives the outcome a for her measurement and $\rho_{Q_B}(a)$ is the marginal state on Bob's system given that Alice received the outcome a.

(a) Write down $p_A(a)$ and $\rho_B(a)$ in terms of $\rho_{Q_A Q_B}$ and $\{M_a\}_a$.

(b) Suppose that $\rho_{Q_A Q_B} = |\Phi^+\rangle\langle\Phi^+|$ with $|\Phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ and that Alice is measuring in the computational basis $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. Compute $H(A)$ and $H_{\min}(A)$.

(c) It turns out that Alice had bought her device from Bob and he told her that she can use it to generate a random bit that is totally secret (no one can guess it with probability more than 1/2). Based on the statistics she observes in her measurements she is convinced. Explain why she is mistaken and how a malicious Bob could obtain her 'secret' random bit.

11. Consider a pure entangled bipartite state $|\psi\rangle_{AB}$. Suppose we perform on system A a rank-one projective measurement $\{P_i\}_{i=1}^d$, i.e., $P_i = |v_i\rangle\langle v_i|$ for some orthonormal basis $\{|v_i\rangle\}_{i=1}^d$ where d is the dimension of the Hilbert space for system A.

(a) What is the post-measurement state after receiving the outcome a?

(b) Show that this post-measurement state is no longer entangled. I.e., a rank-one projective measurement on a subsystem disentangles it with the larger system.

(c) Find a counterexample to part (b) if we allow for projective measurements with larger rank.

12. Let $|\psi\rangle_{AB}$ be a pure bipartite state.

(a) Show that $|\psi\rangle_{AB}$ is entangled iff the marginal states $\rho_A = \mathrm{Tr}_B\left[|\psi\rangle\langle\psi|\right]$ and $\rho_B = \mathrm{Tr}_A\left[|\psi\rangle\langle\psi|\right]$ are mixed states.

(b) Use part (a) or otherwise to argue that because $|\psi\rangle_{AB}$ is pure, A and B cannot be entangled with another system C.

13. Let $|\psi\rangle_{AB}$ be a pure bipartite state.

(a) Show that

$$|\psi\rangle_{AB} \text{ is entangled} \quad \Longleftrightarrow \quad H(A|B) < 0. \tag{45}$$

That is, the conditional entropy of a pure bipartite state is a necessary and sufficient condition for the state to be entangled.

(b) Find a counterexample in the case of mixed states ρ_{AB}. Hint: consider the isotropic states

$$\rho_{AB} = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$$

for $p \in [0, 1]$ where $|\Phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$.

14. Consider a cq-state $\rho_{XB} = p_X(0)|0\rangle\langle 0| \otimes \rho_B(0) + p_X(1)|1\rangle\langle 1| \otimes \rho_B(1)$ where the classical system X is binary. By considering question 11 from the previous week's exercise sheet (or otherwise) establish that

$$H_{\min}(X|B) = -\log\left(\frac{1}{2} + \frac{1}{2}\|p_X(0)\rho_B(0) - p_X(1)\rho_B(1)\|_1\right) \tag{46}$$

15. (Data processing inequality) A very useful concept in (quantum) information theory is that of data processing. Roughly a data processing inequality states that some quantity is monotonic under a (quantum) channel (see question 14 on last week's sheet). For example suppose we have a bipartite system $\rho_{AB} \in \mathcal{D}(AB)$ and a channel $\mathcal{E} : \mathcal{L}(B) \to \mathcal{L}(C)$. If we define $\tau_{AC} = (\mathcal{I} \otimes \mathcal{E})(\rho_{AB})$ then we have a data processing inequality for conditional entropies

$$H(A|B)_{\rho_{AB}} \leq H(A|C)_{\tau_{AC}} \tag{47}$$

and

$$H_{\min}(A|B)_{\rho_{AB}} \leq H_{\min}(A|C)_{\tau_{AC}} \,. \tag{48}$$

We can interpret this roughly as saying that if we hold system B we cannot increase our information about system A just by performing local interactions.[5]

Use the data processing inequality to show that for any tripartite system ABC we have

$$H(A|BC) \leq H(A|B) \tag{49}$$

and

$$H_{\min}(A|BC) \leq H_{\min}(A|B) \,. \tag{50}$$

The first of these inequalities is sometimes known as *strong subadditivity* and captures the notion that by throwing away information we cannot increase our knowledge of a system.

16. Suppose we have a cqq-state $\rho_{XBC} = \sum_x p_X(x)|x\rangle\langle x| \otimes \rho_B(x) \otimes \rho_C$. Note that $\rho_{XBC} = \rho_{XB} \otimes \rho_C$, i.e., system C is independent of X and B.

Show that in this case we have

$$H_{\min}(A|BC) = H_{\min}(A|B) \,. \tag{51}$$

This formalizes our intuition that an independent system does not bring any useful information about A with it.

17. (Concavity[6] of von Neumann entropy) Let $\rho_{AB} = \sum_i q_i \rho_{AB}(i)$ where $q_i \geq 0$ and $\sum_i q_i = 1$ and $\rho_{AB}(i) \in \mathcal{D}(AB)$ – we can ρ_{AB} a *convex mixture* of the states $\rho_{AB}(i)$. Prove that

$$H(A|B)_{\rho_{AB}} \geq \sum_i q_i H(A|B)_{\rho_{AB}(i)} \,. \tag{52}$$

[Hint: it might be worth considering a qqc-state ρ_{ABX} such that $\rho_{AB} = \mathrm{Tr}_X\left[\rho_{ABX}\right]$.]

18. Show that any separable state ρ_{AB} (not entangled) satisfies

$$H(A|B)_{\rho_{AB}} \geq 0 \,. \tag{53}$$

I.e., negative entropy is a sufficient condition for a bipartite state to be entangled. However, by part (b) of question 14 it is not necessary. [Hint: consider the previous question.]

Show that this also implies that $H(X|B) \geq 0$ for any cq-state ρ_{XB}.

[5] It seems that any useful measure of information should obey such a data processing inequality.

[6] Recall that a function f is called concave if $f(\sum_i q_i x_i) \geq \sum_i q_i f(x_i)$ where $q_i \geq 0$ and $\sum_i q_i = 1$.