University of California at San Diego – Department of Physics – Prof. John McGreevy

Physics 213/113 Winter 2023
Assignment 3 – Solutions

Due 11:00am Tuesday, January 31, 2023

1. **Chain rules.** [optional]

   Show that for a joint distribution of n random variables $p(X_1 \cdots X_n)$, the joint and conditional entropies satisfy the following chain rule:

   $$H(X_1 \cdots X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1} \cdots X_1).$$

   Show that the $n = 2$ case is the expectation of the log of the BHS of Bayes rule. Then repeatedly apply the $n = 2$ case to increasing values of n.

2. **Learning decreases ignorance only on average.**

   Consider the joint distribution $p_{yx} = \begin{pmatrix} 0 & a \\ b & b \end{pmatrix}_{yx}$, where $y =\uparrow,\downarrow$ is the row index and $x =\uparrow,\downarrow$ is the column index (so yx are like the indices on a matrix). Normalization implies $\sum_{xy} p_{xy} = a + 2b = 1$, so we have a one-parameter family of distributions, labelled by b.

   What is the allowed range of b?

   Find the marginals for x and y. Find the conditional probabilities $p(x|y)$ and $p(y|x)$.

   Check that $H(X|Y) \leq H(X)$ and $H(Y|X) \leq H(Y)$ for any choice of b.

   Show, however, that $H(X|Y =\downarrow) > H(X)$ for any $b < \frac{1}{2}$.

   Since $a = 1 - 2b$, we need $0 \leq b \leq \frac{1}{2}$ to keep all the probabilities positive. The marginals are $p_x = (b, a + b)$ and $p_y = (a, 2b)$. The conditional probabilities are

   $$p(x|y) = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}_{yx}, \quad p(y|x) = \begin{pmatrix} 0 & 1 \\ \frac{a}{a+b} & \frac{b}{a+b} \end{pmatrix}_{yx}.$$

   So we have $H(X) = -b \log b - (a + b) \log(a + b) = -b \log b - (1 - b) \log(1 - b) = H(b, 1 - b)$. So if we measure Y and get $\downarrow$, we know nothing about X. This means that for this particular outcome

   $$H(X|y =\downarrow) = 1 > H(X)$$

But if we get $y =\uparrow$ then we know for sure $X =\downarrow$: $H(X|y=\uparrow) = 0$, no uncertainty. The *average* conditional entropy for X is then

$$H(X|Y) \equiv \langle H(X|Y) \rangle_{XY} = p_{y=\downarrow} H(X|y=\downarrow) + p_{y=\uparrow} H(X|y=\uparrow) = 2b.$$

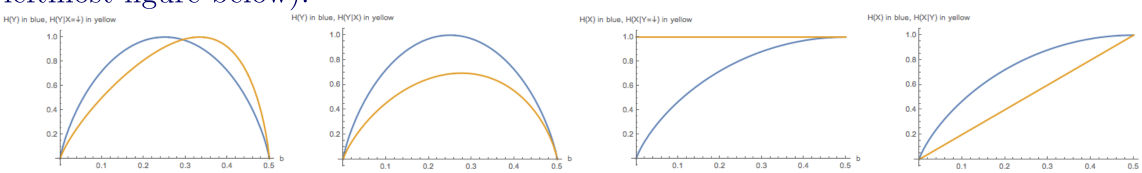
And indeed and $H(b, 1-b) > 2b$ for $b < \frac{1}{2}$, as you can see in the rightmost figure below. Similarly,

$$H(Y|x=\uparrow) = 0$$

means that the uncertainty for $x =\downarrow$ must be larger. Specifically,

$$H(Y) = H(1-2b) \geq H(Y|X) = (1-b) H\left(\frac{1-2b}{1-b}\right)$$

but $H(Y|X=\downarrow) = H\left(\frac{1-2b}{1-b}\right)$ can be either larger or smaller than $H(Y)$ (as in the leftmost figure below).



3. **Mutual information bounds correlations.** Consider again the distribution on two binary variables from the last problem: $p_{yx} = \begin{pmatrix} 0 & a \\ b & b \end{pmatrix}_{yx}$, where $y = 1, -1$ is the row index and $x = 1, -1$ is the column index (so yx are like the indices on a matrix). Normalization implies $\sum_{xy} p_{xy} = a + 2b = 1$, so we have a one-parameter family of distributions, labelled by b.

   (a) I've changed the labels on the variables from $\uparrow, \downarrow$ to $1, -1$ so that we can consider correlation functions, such as the connected two-point function

   $$C \equiv \langle xy \rangle_c \equiv \langle xy \rangle - \langle x \rangle \langle y \rangle$$

   where $\langle A \rangle \equiv \sum_{xy} p_{yx} A$. Compute C as a function of b.

   $$C = \sum_{xy} p_{xy} xy - \left(\sum_x p_x x\right)\left(\sum_y p_y y\right) = -a + a(a-2b)|_{a=1-2b} = 4b(2b-1).$$

   (b) Compute the mutual information between X and Y

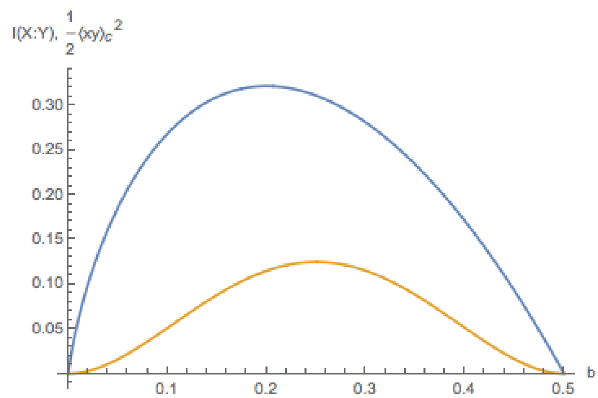
   $$I(X:Y) = \sum_{xy} p_{yx} \log \frac{p_{yx}}{p_y p_x}.$$

$$I(X:Y) = \sum_{xy} p_{xy} \log \frac{p_{xy}}{p_x p_y} = \left(0 + a \log \frac{a}{a(a+b)} + b \log \frac{b}{2b^2} + b \log \frac{b}{(a+b)2b} \right) |_{a=1-2b}.$$

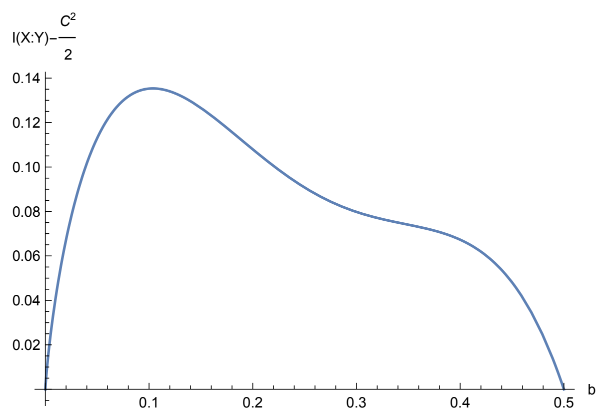
(c) Check that

$$I(X:Y) \geq \frac{1}{2} C^2$$

for every value of b (for example, plot both functions).



Actually, even nicer is to plot the difference (I am using natural log here)



(d) [Bonus] The inequality I quoted in lecture, and which we will prove in the more general quantum case later, is

$$I(X:Y) \geq \frac{1}{2} \frac{\langle \mathcal{O}_X \mathcal{O}_Y \rangle_c^2}{\| \mathcal{O}_X \|^2 \| \mathcal{O}_Y \|}$$

where the norms are defined (in the classical case) by

$$\| \mathcal{O}_X \|^2 \equiv \sup_{p | \sum_x p_x = 1} \{ \sum_x \mathcal{O}_x^\star \mathcal{O}_x p_x \}.$$

Show that in the above example, the 'operators' x, y are normalized, in the sense that $\| x \| = \| y \| = 1$.

The functions were ± 1, so their maximum absolute squares are just 1.

4. **Strong subadditivity, the classical case.** [From Barnett] Prove *strong subaddivity* of the Shannon entropy: for any distribution on three random variables,

$$H(ABC) + H(B) \leq H(AB) + H(BC).$$

(The corresponding statement about the von Neumann entropy is not quite so easy to show.)

Hint: $q(a,b,c) \equiv \frac{p(a,b)p(b,c)}{p(b)}$ is a perfectly cromulent probability distribution on ABC.

What is the name for the situation when equality holds? Write the condition for equality in terms of the conditional mutual information $I(A : C|B)$.

The relative entropy is positive:

$$\begin{aligned} 0 \leq D(p(ABC)||q(ABC)) &= \sum_{abc} p(abc) \left(\log p(abc) - \log p(ab)p(bc) + \log p(b) \right) \\ &= -H(ABC) - \sum_{ab} p(ab) \log p(ab) - \sum_{bc} \log p(bc) + \sum_{b} \log p(b) \\ &= -H(ABC) + H(AB) + H(BC) - H(B). \end{aligned}$$

The ease with which that just happened contrasts stiffly with the quantum case.

5. **Symbol coding problem.** [Important] You are a mad scientist, but a sloppy one. You have 127 identical-looking jars of liquid, and you have forgotten which one is the poison one. You have at your disposal 7 rats on whom your poor moral compass will allow you to test the liquids. However (the rats have a strong social network and excellent spies) you only get one shot: the rats must drink all at once (or they will catch on to what is happening and revolt). You may mix the liquids in separate containers. Any rat that drinks any amount of poison will turn bright orange. Design a protocol to uniquely identify the poison jar.

Number the jars 1 to 127, and write these numbers in binary, so $1 = 0000001, 2 = 00000010, ...127 = 1111111$. Number the rats 1 to 7. Feed a little bit of the liquid from jar number $N = s_1..s_7$ to the rats i with $s_i \neq 0$. Line up the rats and read off the binary digits according to orange = 1, not orange = 0. This is the number of the jar with the poison.

Another way to describe this solution, which may be more intuitive is as binary search: divide the poison jars into two groups. Feed a mixture of the first half to the first rat. If he turns orange, repeat this procedure with just the first half of the jars and the second rat. Repeat until the jar is uniquely identified. Since $128 = 2^7$, we need at most 7 steps.

6. **Another coding problem.** [optional, but how can you resist?] The problem is to establish a code by which you can transmit to your friend a number from $1 \cdots N = 64$. The tool you will be given is a chessboard (8×8) where an adversary has randomly placed identical markers on some of the squares. You are only allowed to add or remove a single marker. Your friend will see only the result of your actions, not the initial configuration. You may speak to your friend beforehand.

   I learned this problem from A. Manohar.

   I think I should not post the solution because it may deprive someone of the satisfaction of solving this nice puzzle. If you really want me to tell you the answer, come talk to me.