1. Mutual information bounds correlations. Consider again the distribution on two binary variables from last homework: \( p_{yx} = \begin{pmatrix} 0 & a \\ b & b \end{pmatrix} \), 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 \( b \).

(b) Compute the mutual information between \( X \) and \( Y \)

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

(c) Check that

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

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

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

\[
I(X : Y) \geq \frac{1}{2} \langle O_X O_Y \rangle_c^2 \leq \frac{\sum_{x} O_x^2 O_x p_x}{\langle O_X \rangle \langle O_Y \rangle}
\]

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

\[
\| O_x \|^2 \equiv \sup_{p} \left\{ \sum_{x} O_x^2 O_x p_x \right\}.
\]

Show that in the above example, the ‘operators’ \( x, y \) are normalized, in the sense that \( \| x \| = \| y \| = 1 \).
2. **Strong subadditivity, the classical case.** [From Barnett] Prove strong subadditivity 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 \( I(A : C | B) \).

3. **Symbol coding problem.** 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.

4. **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 \((8 \times 8)\) 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.

5. **Huffman code.** Make the Huffman code for the probability distribution \( p(x) = \{ .5, .2, .15, .1, .05 \} \).

Compare the average word length to the Shannon entropy.

Bonus: what property of the distribution determines the deviation from optimality?

6. **Huffman code decryption problem.** [Optional, but fun.]

0010 1100010000101011 0101010100101110 00011001000011 0011010011 10101001000010001110101101101101 1111100011 2
7. **Analogy with strong-disorder RG.** [open ended, more optional question]

Test or decide the following consequence suggested by the analogy between Huffman coding and strong-disorder RG: The optimality of the Huffman code is better when the distribution is broader. A special case is the claim that the Huffman code is worst when all the probabilities are the same. Note that the outcome of the Huffman algorithm in this case depends on the number of elements of the alphabet.

Measure the optimality by $\langle \ell \rangle - H[p]$ (or maybe $\frac{\langle \ell \rangle - H[p]}{H[p]}$?).

8. **Binary symmetric channel.**

For the binary symmetric channel $ABE$ defined in lecture, with $a, b, e \in \{0,1\}$, and

$$p(a) = (p, 1-p)_a, \quad p(e) = (q, 1-q)_e,$$
and $b = (a+e)_2$,

find all the quantities $p(a, b), p(b|a), p(a|b)$ and $H(B), H(B|A), I(B : A), I(B : A|E)$. Find the channel capacity.