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

Physics 239/139 Fall 2019
Assignment 4 – Solutions

Due 11:00am Tuesday, February 7, 2023

1. **Error correcting code brain-warmer.**

   (a) For the [7,4] Hamming code discussed in lecture, check that $Ht = 0$ where t is any codeword, and H is the given parity check matrix.

   A codeword is $t = Gs$, so $Ht = 0$ follows from

$$HG = \begin{pmatrix} 1&1&1&0&1& & \\ 1&1&0&1& &1& \\ 0&1&1&1& & &1 \end{pmatrix} \begin{pmatrix} 1& & & \\ &1& & \\ & &1& \\ & & &1 \\ 1&1&1&0 \\ 1&1&0&1 \\ 0&1&1&1 \end{pmatrix} = 0.$$

   Modulo two, of course!

   (b) If you (B) are communicating with someone (A) through a channel with $H(A|B) = 1/7$ using this code and you receive the string

$$r = (0, 1, 0, 0, 0, 0, 1)^t,$$

   what is the most likely intended message string?

The given $H(A|B)$ says that we should expect one error every 7 bits, so a single wrong bit. The syndrome is $z = Hr = (1, 1, 0)$, which means that parity checks 5 and 6 fail, and 7 succeeds. Referring to the figure at right (this is the convention in my notes; in lecture I swapped 6 and 7), this is the syndrome for a single error in message bit 1. So the most likely message is obtained by flipping the first bit of the received signal:

$$s_{\text{most likely}} = (1, 1, 0, 0).$$

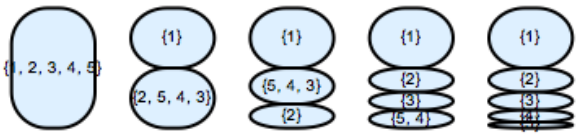
The received message could also have been obtained by more errors from a different intended message, but this is less likely.

2. **Huffman code.** Make the Huffman code for the probability distribution $p(x) = \{.5, .2, .16, .1, .04\}$.

   Compare the average word length to the Shannon entropy.

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

   Using the coarse-graining steps (time goes to the left) :



   I find the codewords `0, 11, 100, 1010, 1011`. My conventions are: the less probable element gets the 1, I sort the list at each step, if there's a tie I do not switch the order. Different conventions will lead to different codewords. The Shannon entropy is 1.92322 and the average code length is 1.95.

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

   `0101101000011010100100100001111011011010000111001001000011100 0,`
   `010101010010110 110000000001000011 10101010010110001100011 1111100011`
   `11110010110000100010110100 1011. 001010101 010101010010110`
   `10110100001100011 0000 000110010101011010101110010110011111`

```
111011000100100111 1010101001001 0101010100101101001
1101011111110111001
10101100101100010010111011001100110101100100111000, 1000010010111011
010010101 1111100011 11010111111110111001100 101001001011011101
01011010010111011 010010110111 10100110010100001101010O.
```

You might want to use `Mathematica` to do this problem!

Hint: I used the letter frequencies from *The Origin of Species*, which is built into `Mathematica`.

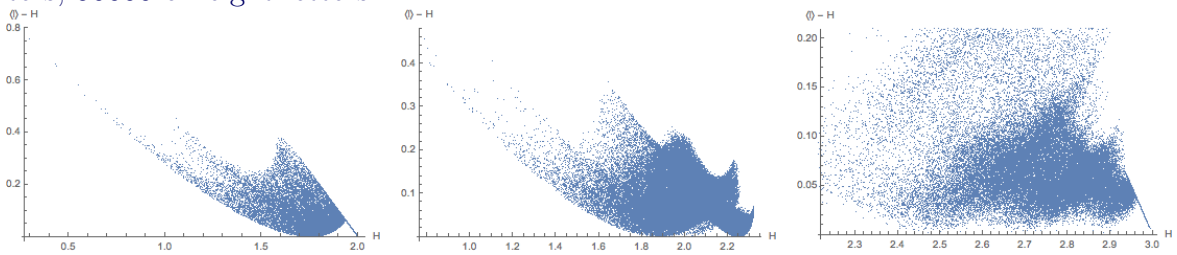
congratulations, you have found the treasure. if you used a different text for your letter frequencies, some of the letters will come out wrong.

4. **Analogy with strong-disorder RG.** [open ended, more optional question]

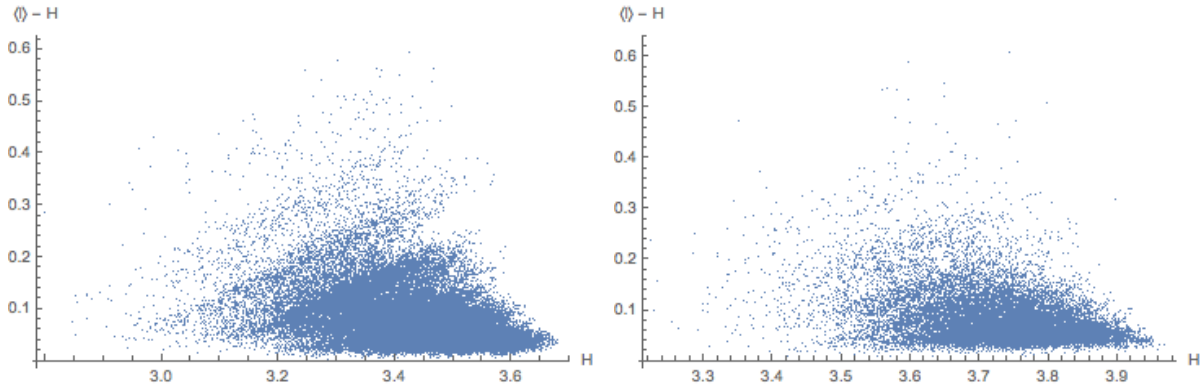
   Test or decide the following consequence suggested by the analogy between Huffman coding and strong-disorder RG: The optimality of the Huffman code is better when the distribution is broader. A special case is the claim that the Huffman code is worst when all the probabilities are the same. Note that 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 $\frac{\langle \ell \rangle - H[p]}{H[p]}$?).

   It is possible to make some pretty pictures this way. I find that the general trend is in the direction predicted by the SDRG intuition, but with lots of number-theoretic features which deviate from it. Here is deviation from optimal average codeword length for 20000 random distributions on four letters, 10^5 on five letters, 50000 on eight letters:



   and 20000 on 13 and 16 letters each:



Powers of two are special because then it is possible to saturate the bound with a Huffman code, as in the initial Shannon example I showed in lecture. This was a successful problem if only in that these pictures look like murmurations.

5. **Maximum entropy with fixed average.** Suppose we have a probability distribution π_ℓ on a positive integer ℓ with fixed average:

$$\langle \ell \rangle \equiv \sum_{\ell=1}^{\infty} \pi_\ell \ell = a. \tag{1}$$

Bound the Shannon entropy of π from above:

$$H[\pi] \leq a \log a - (a-1) \log(a-1) \ . \tag{2}$$

Show that this goes like $\log a$ at large a.

[Hint: use Lagrange multipliers to maximize the entropy subject to the normalization condition and the condition on the average.]

We'll maximize the function

$$I[\pi] = H[\pi] + \beta(\sum_{\ell=1}^{\infty} \ell\pi_\ell) + \lambda(\sum_{\ell=1}^{\infty} \pi_\ell - 1). \tag{3}$$

The derivative gives

$$0 = \frac{\partial I}{\partial \pi_\ell} = \log \pi_\ell - 1 + \lambda + \lambda\ell \tag{4}$$

which is solved by

$$\pi_\ell = \mathcal{N} e^{-\beta\ell}. \tag{5}$$

The two Lagrange multipliers will be happy if

$$1 = \sum_{\ell=1}^{\infty} \pi_\ell = \mathcal{N} \sum_{\ell=1}^{\infty} e^{-\beta\ell} = \frac{\mathcal{N}}{e^\lambda - 1} \quad \Longrightarrow \mathcal{N} = e^\lambda - 1 \tag{6}$$

and

$$a = \sum_{\ell=1}^{\infty} \ell \pi_\ell = \mathcal{N} \sum_{\ell=1}^{\infty} \ell e^{-\beta \ell} = \frac{\mathcal{N} e^{\lambda}}{(e^{\lambda}-1)^2} \quad \Longrightarrow \; e^{\lambda} = \frac{a}{a-1}. \tag{7}$$

Thus the max-entropy distribution with fixed average is

$$\pi^{\star}_{\ell} = \left(\frac{a-1}{a}\right)^{\ell} \frac{1}{a-1}. \tag{8}$$

Its entropy is

$$H(\pi^{\star}) = -\log \frac{a-1}{a} \underbrace{\frac{1}{a-1} \sum_{\ell=1}^{\infty} \left(\frac{a-1}{a}\right)^{\ell} \ell}_{=a} + \log(a-1) \underbrace{\frac{1}{a-1} \sum_{\ell=1}^{\infty} \left(\frac{a-1}{a}\right)^{\ell}}_{=1} = a \log a - (a-1) \log(a-1) \tag{9}$$

Taylor expanding about $a = \infty$ (for example by `Series[`$H, \{a, \infty, 1\}$`]` in Mathematica) gives

$$H(\pi^{\star}) \overset{a \to \infty}{\simeq} \log a + 1 - \frac{1}{2a} + \cdots \tag{10}$$

6. **Binary symmetric channel.**

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

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

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

We have:

$$P(b) = \sum_{ae} p(a) q(e) \delta_{b,(a+e)_2}.$$

where $p(a) = (p, 1-p)_a, q(e) = (1-q, q)_e$.

$$P(b=0) = p(1-q) + (1-p)q, P(b=1) = pq + (1-p)(1-q).$$

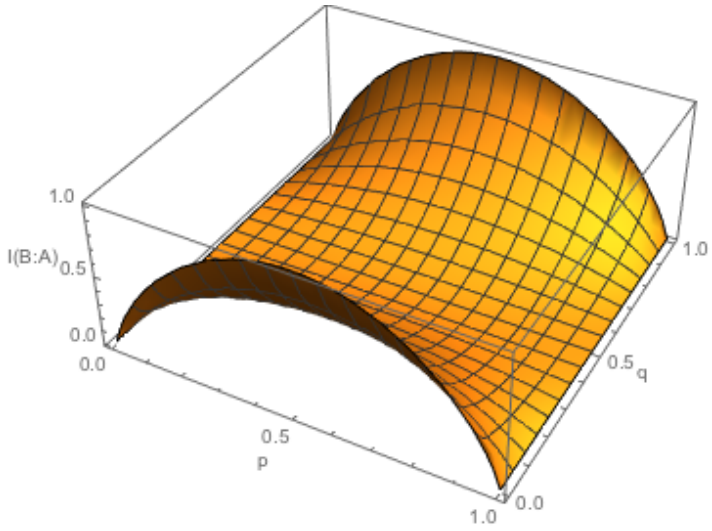
$$H(B) = H_2(p(1-q) + (1-p)q).$$

$$p(b|a) = \begin{pmatrix} 1-q & q \\ q & 1-q \end{pmatrix}_{ab}$$

$$\begin{aligned} H(B|A) &= \sum_a p(a) H(B|A=a) = \sum_a p(a) \left(- \sum_b p(b) \log p(b) \right) \\ &= p \left(H_2(q) \right) + (1-p) \underbrace{H_2(q)}_{= H_2(1-q)} = H_2(q). \end{aligned} \tag{11}$$

$$I(B:A) = H(B) - H(B|A) = H_2(p(1-q) + (1-p)q) - H_2(q).$$

The channel capacity is obtained by maximizing this function of p (for each fixed q):



which happens at $p = 1/2$, since

$$\partial_p I(B:A) = (2q-1)\log\left(-1 - \frac{1}{-1+p+q-2pq}\right)$$

which vanishes for generic q only if the argument of the log is

$$1 = -1 - \frac{1}{-1+p+q-2pq} \quad \leftrightarrow \quad p = \frac{1}{2}.$$

The value at the maximum is:

$$C(q) = 1 - H_2(q).$$

To find $I(B:A|E)$, I found it easiest to do the following:

$$I(B:A|E) = H(B|E) - \underbrace{H(B|AE)}_{=0}.$$

Then to find $H(B|E)$, use:

$$p(abe) = \delta_{b,a+e} p(a) q(e)$$

$$p(be) = \sum_a p(abe) = \left(p\delta_{b,e} + (1-p)\delta_{b,e+1}\right) q(e)$$

$$p(b|e) = \frac{p(be)}{p(e)} = p\delta_{b,e} + (1-p)\delta_{b,e+1}.$$

This distribution on b has Shannon entropy $H_2(p)$. Therefore

$$I(B:A|E) = H(B|E) - \underbrace{H(B|AE)}_{=0} = H_2(p)$$

as claimed in lecture.

7. **Chain rule for mutual information.** [optional]

Show from the definitions that the mutual information satisfies the following chain rule:

$$I(X:YZ) = I(X:Y) + I(X:Z|Y) = I(X:Z) + I(X:Y|Z).$$

More generally,

$$I(X_1 \cdots X_n : Y) = \sum_{i=1}^{n} I(X_i Y | X_{i-1} \cdots X_1). \tag{12}$$

The mutual information is $I(X:Y) = H(X) - H(X|Y)$, so the LHS is

$$I(X_1 \cdots X_n : Y) = H(X_1 \cdots X_n) - H(X_1 \cdots X_n|Y) \tag{13}$$

$$= \sum_{i=1}^{n} H(X_i|X_{i-1} \cdots X_1) - \sum_{i=1}^{n} H(X_i|X_{i-1} \cdots X_1 Y). \tag{14}$$

In the second step we used the chain rule for the Shannon entropy. The RHS is

$$\sum_{i=1}^{n} I(X_i Y|X_{i-1} \cdots X_i) = \sum_{i=1}^{n} \left(H(X_i|X_{i-1} \cdots X_1) - H(X_i|X_{i-1} \cdots X_1 Y) \right),$$

which is the same expression. To get the first two equalities, take $n = 3$ and $X_1 = Y, X_2 = Z, Y = X$ and $X_1 = Z, X_2 = Y, Y = X$ respectively in (12), and use liberally the fact that the mutual info is symmetric in the sense that $I(A:B) = I(B:A)$.

Alternatively, we can use the definition more directly, eg in the case $n = 2$:

$$I(X:Z|Y) = \sum_{xyz} p(y)p(xz|y) \log \frac{p(xz|y)}{p(x|y)p(z|y)} \tag{15}$$

$$\overset{\text{Bayes}}{=} \sum_{xyz} p(xyz) \log \frac{p(xzy)p(y)}{p(xy)p(zy)} \tag{16}$$

$$= \sum_{xyz} p(xyz) \log \frac{p(xyz)}{p(yz)p(x)} + \sum_{xyz} p(xyz) \log \frac{p(x)p(y)}{p(xy)} \tag{17}$$

$$= I(X:YZ) - I(X:Y). \tag{18}$$

8. **Measuring entropy.** [optional] Build a Lempel-Ziv encoder (and a decoder, so you can check that it works), and use it to estimate the entropy of some stream of data.

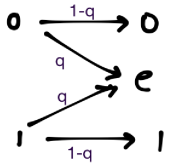
   Where to get the data: you could use a string of iid bits (in which case it's easy to check your answer), or you could use the output of a model of some physical process. A good example is the one in section IIA of this paper:

> N particles are distributed on a chain of $L > N$ sites with no multiple occupancy allowed. At each time step, represent the occupation numbers of particles as a string of L 0s and 1s. A site is considered *active* if one of its neighbors is also occupied. At each time step, randomly select an active particle and move it to an unoccupied site. A useful order parameter is the fraction f of active sites. If $f \to 0$, the dynamics stops. There is a critical value of the density $\rho = N/L$ above which the system never reaches an $f = 0$ state. Can you see the critical value of f in the entropy?

   This reference also explains an implementation of the LZ77 algorithm.

9. **Binary erasure channel.**

   Find the channel capacity of this channel:



   The channel is defined by

$$p(b|a) = \begin{pmatrix} 1-q & 0 \\ 0 & 1-q \\ q & q \end{pmatrix}_{b=0e1, a=01} . \tag{19}$$

   Let $p(a=0) \equiv p$ determine the input distribution. Then the joint distribution is

$$p(ab) = \big(p(b|a)p(a)\big)_{a,b} = \begin{pmatrix} p(1-q) & pq & 0 \\ 0 & (1-p)q & (1-p)(1-q) \end{pmatrix}_{a=01, b=0e1}$$

   This has the same form as the binary symmetric channel with some extra zeros. The distribution for the output variable is $p(b) = \big(p(1-q)\ q\ (1-p)(1-q)\big)_b$. We find for the mutual information

$$I(A:B) = H(A) + H(B) - H(AB) = H_2(p)(1-q)$$

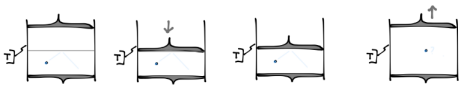
which is maximized when $p = \frac{1}{2}$, and therefore the capacity is

$$\max_{p(A)} I(A : B) = 1 - q.$$

Satisfyingly, this is also the probability that any bit we send is not erased.

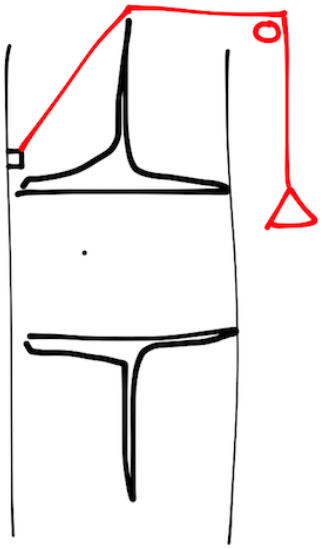
10. **Mechanical engineering problem.** [optional]

In lecture I claimed that the expansion of an ideal gas against a piston, as in the figure at right, could be used to lift a weight.



Design a plausible system of strings and pulleys to make this happen.

Here's a nice solution, from Akhil Premkumar:



11. **Test of Landauer's Principle.** [optional]

Consider logical bits which are stored in the magnetization (up or down) of little magnets. Show that copying a known bit (say 0) onto an *unkown* bit by the method described in lecture costs energy at least $k_B T \ln 2$.

Clearly this process cannot be reversible, since the initial state of the target bit is lost in the process. In the physical realization of the bit as a column of gas, it is clear that we must do work $k_B T \ln 2$ to accomplish this. I was hoping someone would show how this constraint is manifested in the realization in terms of the deformable double-well potential. In particular, it must be that a protocol for deforming the potential which is guaranteed to adiabatically take the target bit from one of the two states (say 0) onto the state of the given bit to be copied must fail if employed with the target bit initialized in the other state (1).