Table of Contents

Directions for homework submission

Submit each of your homework to canvas as the pdf output of the jupyter notebook, and the jupyter notebook (.ipynb). Name your files starting in the format of "Last_name_First_name_File_name" separated by underscores.

For example, Jieyu submits two files for her homework this week:

1. Zheng_Jieyu_HW1.pdf (the pdf output of the jupyter notebook)

   If you have problems rendering your notebook into pdf, you can open your notebook in a browser and print -> save as pdf.
2. Zheng_Jieyu_HW1.ipynb

Please make sure your notebook can be run without errors within the cns187 virtual environment. Any file that fails to be executed on TA's end will be considered as late submissions.

Caltech Honor code: Searching for the solutions online is strictly prohibited. You should refer to the textbooks and lecture slides. If you are citing any external sources online, please include a list of references.

Collaboration on homework assignments is encouraged. However, **you cannot show each other the numerical answers or codes**. Please note at the beginning of each answer whom you have discussed the problems with (including TAs).

All the mathematics should be typed in Latex format. You may work on a piece of paper and then type it into the notebook. Here is a useful cheat sheet (http://users.dickinson.edu/~richesod/latex/latexcheatsheet.pdf). Please do not submit pictures of handwritten maths.

For the schematic and drawings to be submitted, please display the images in markdown cells in your homework amd make sure they show up in your pdf rendering.

Please make sure that all your plots include a title and axis labels with units. One point will be deducted for each missing element.

```python
In [1]:  import numpy as np
         import matplotlib.pyplot as plt
```

Mid-term evaluation form (5 pt)

Please complete the mid-term evalutation form (https://forms.gle/dd3HnM8o78jZd76J6).

Bonus points (5 pt): Leave a few comments for the optional questions!

Done

Order of magnitude neuroscience: The information rate of human behaviors (20 pt)

Calculate the information rate of the following three behaviors based on the given references:

1. World record in Speedcard (memorization) (https://en.wikipedia.org/wiki/World_Memory_Championships#Records)
2. World record in solving Rubik's cube (https://www.youtube.com/watch?v=3UTQXc-nXOM)
3. Action per minute of a professional e-athlete (https://en.wikipedia.org/wiki/Actions_per_minute)

Hint: Recall that the information is the difference between two entropies. What is the uncertainty about the person's actions before the behavioral task? And how much is that reduced by observing the behavior? List all the other assumptions that you have made for each case, and be sure to include additional references should you need them.

What is common across these numbers?

For this problem let $IR :=$ information rate

1. To calculate entropy we consider the total number of possible card permutations from which the single shuffle is chosen for a given speedcard game, and this gives:

$$H(x) = \log_2(52!) = 225.58 \text{ bits}$$

Additionaly the world record is 12.74 s, so we have:

$$IR = \frac{225.58 \text{ bits}}{12.74 \text{ s}}$$

$$IR = 17.7064 \; \frac{\text{bits}}{\text{s}}$$

2. To calculate entropy we consider the total number of permutations that exist for a Rubik's cube, and this gives:

$$H(x) = \log_2(43, 252, 003, 274, 489, 856, 000) = 65.23 \text{ bits}$$

Additionally the world record is 15.5 s, so we have:

$$IR = \frac{65.23 \text{ bits}}{15.5 \text{ s}}$$

$$IR = 4.20839 \; \frac{\text{bits}}{\text{s}}$$

This is operating under the assumption that a person actually considers all permutations while taking the time to solve a Rubik's cube. In reality, solving Rubik's cubes typically relies on algorithms more so as opposed to an actual process of storing the permutations as bits in the brain.

3. We will assume that any given action takes N bits, although in reality, the different actions may use different amounts of information capacity. The noted record for APM is $818 \frac{\text{actions}}{\text{min}}$. We can convert this to $\frac{\text{bits}}{\text{s}}$:

$$\frac{818 \text{ actions}}{1 \text{ min}} \times \frac{1 \text{ min}}{60 \text{ s}} \times \frac{N \text{ bits}}{1 \text{ action}} = N \cdot \frac{818 \text{ bits}}{60 \text{ s}}$$

So our information rate is as so:

$$N \cdot 13.6333 \; \frac{\text{bits}}{\text{s}}$$

The commonality among all these rates are that they are on the order of 10 bits/sec which is around the calculated information throughput of humans as shown in lecture.

The entropy of a discrete source (20 pt)

Say an information source uses symbols $i = 1, \ldots, n$ with probabilities $p_1, \ldots, p_n$. Show that the entropy of the source is largest if all the symbols occur with equal probability, i.e. $p_i = 1/n$ for all n.

We make use of Jensen's inequality which states that for any concave down function f and discrete random variable $\mathbf{X}$:

$$f\left(\mathbf{E}\left[\mathbf{X}\right]\right) \geq \mathbf{E}\left[f\left(\mathbf{X}\right)\right]$$

In this case we consider $f\left(\mathbf{X}\right) = \log_2\left(\mathbf{X}\right)$.

Let
$\mathbf{S} :=$ R.V. representing the source that takes on N possible values with probabilities p

Let $p\left(\mathbf{S}\right) := p.m.f.$ of $\mathbf{S}$.

$$H(x) = -\sum_{j=1}^{N} p_j \cdot \log_2\left(p_j\right)$$

$$H(x) = -\mathbf{E}\left[\log_2\left(p\left(\mathbf{S}\right)\right)\right]$$

$$H(x) = \mathbf{E}\left[\log_2\left(\frac{1}{p\left(\mathbf{S}\right)}\right)\right] \leq \log_2\left(\mathbf{E}\left[\frac{1}{p\left(\mathbf{S}\right)}\right]\right)$$

$$H(x) \leq \log_2\left(\sum_{j=1}^{N} p_j \cdot \frac{1}{p_j}\right) = \log_2\left(\sum_{j=1}^{N} 1\right)$$

$$H(x) \leq \log_2(N)$$

This shows the maximum entropy is achieved when $p_i = \frac{1}{N}$ for all N since the entropy for such a case is exactly $\log_2(N)$

Diagnostic testing (25 pt)

Suppose a rare disease afflicts individuals in a population with relative frequency 0.001. You have available a diagnostic test for this disease. Unfortunately, the test is not perfect, but occasionally produces false positive results or false negative results, and both happen to occur with a probability of 0.001. You perform this test on a randomly chosen subject and obtain a positive result.

a) What was your uncertainty (in bits) about the subject's disease status before performing the test?

$$\underline{x: \text{Disease status}}$$

$$x_1 = \text{event that a person has the disease}$$

$$x_2 = \text{event that a person doesn't have the disease}$$

$$\mathbf{P}(x_1) = 0.001 \quad \mathbf{P}(x_2) = 0.999$$

$$H(x) = -\mathbf{P}(x_1) \cdot \log_2 \mathbf{P}(x_1) - \mathbf{P}(x_2) \cdot \log_2 \mathbf{P}(x_2)$$

$$H(x) = -0.001 \cdot \log_2(0.001) - 0.999 \cdot \log_2(0.999) \approx 0.01140775774$$

$$H(x) \approx 0.01141 \text{ bits}$$

b) What is the uncertainty after seeing the test result?

$$\underline{\text{y: Patient has disease}}$$

$$\mathbf{P}(y) = \mathbf{P}(y \cap x_1) + \mathbf{P}(y \cap x_2) = \mathbf{P}(x_1) \cdot \mathbf{P}(y|x_1) + \mathbf{P}(x_2) \cdot \mathbf{P}(y|x_2)$$

$$\mathbf{P}(y) = 2 \cdot (0.001 \cdot 0.999) = 0.001998$$

$$\mathbf{P}(y|x_1) = 0.999$$

$$\mathbf{P}(y|x_2) = 0.001$$

$$\mathbf{P}(x_1|y) = \frac{\mathbf{P}(y|x_1) \cdot \mathbf{P}(x_1)}{\mathbf{P}(y)} = 0.5$$

$$\mathbf{P}(x_2|y) = \frac{\mathbf{P}(y|x_2) \cdot \mathbf{P}(x_2)}{\mathbf{P}(y)} = 0.5$$

$$H(x|y) = -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5) = 1 \text{ bit}$$

c) How much information have you gained from this specific test outcome?

$$I(x; y) = H(x) - H(x|y) \approx -0.98859 \text{ bits}$$

d) How do you reconcile this with Prof Meister's claim that the mutual information is always positive?

The negative value for mutual information is due to the test result providing misinformation about the patient's disease status. This is more than likely a result of the fact that the probability of having the disease is so low that it is hard to gain reasonable information on disease status with the given rates of false positives and false negatives. The mutual information, in theory, is always positive given that it is a measure of the reduction in uncertainty. However, it seems that in practice, due to limitations in the accuracy of the measurement/in the model used to estimate the probabilities, the resultant mutual information can be negative, as seen above.

e) What should you do next? [3 sentences]

For some health policy implications of this, see e.g. Hoffrage, U., S. Lindsey, R. Hertwig, and G. Gigerenzer. 2000. Medicine. Communicating statistical information. Science. 290:2261-2. DOI: 10.1126/science.290.5500.2261

Given that the positive test result provided negative information or misinformation about the subject's disease status, further evaluation would be required to confirm the presence of the disease. Other methods should be explored to avoid misdiagonosing the patient. Oftentimes with tests such as this, falst positives occur, so certainty should be increased before reliably diagonosing. In this case, the calculation shows certainty to have decreased.

Limitation of information gain (30 pt)

From the lectures, we learned that single neurons can encode subtle changes of sensory inputs with surprising fidelity. They sometime overperform the perceptual sensitivity of trained psychophysical observers.

This might be counter-intuitive. Pooling responses across neurons should have averaged out noise in the activity of single cells and lead to substantially improved psychophysical performance than single cells.

a) Based on Population Code Lec 2, what is one possible reason that pooling responses from the neuron ensemble does not improve overall psychophysical performance?

This could be a result of noise correlations between the neurons being tracked that causes limitations to be imposed on information gain from population decoding.

b) Assuming we have a population of m artificial cells with simplified signal response with identical distribution, let's examine how the correlation within the population will affect the overall signal-to-noise ratio (SNR).

In the present context, each signal distribution represents the responses of a single neuron. We will plot the signal-to-noise ratio of M summed signals, $X_1 \ldots X_m$.

We define the sum of the M signals as:

$$Y = \sum_i^m X_i$$

we define the signal as the expectation of Y, and the noise as the standard deviation of Y.

$$S = \mathrm{E}[Y]$$
$$N = \sqrt{\mathrm{Var}[Y]}$$

Please derive an expression for S, assuming that X_i has the same mean $\langle X \rangle$. Please also derive an expression for N, assuming that the X_i have pairwise correlation coefficient of r, which means that:

$$\mathrm{Var}[X_i] = \left\langle X_i^2 \right\rangle - \langle X_i \rangle^2 = \sigma^2$$
$$\mathrm{Cov}[X_i, X_j] = \left\langle X_i X_j \right\rangle - \langle X_i \rangle \left\langle X_j \right\rangle = r \cdot \sigma^2$$

$$S = \mathbf{E}\left[\mathbf{Y}\right] = \sum_{i=0}^{M} \mathbf{E}\left[\mathbf{X}_i\right] = \sum_{i=0}^{M} \langle \mathbf{X} \rangle = M \cdot \langle \mathbf{X} \rangle$$
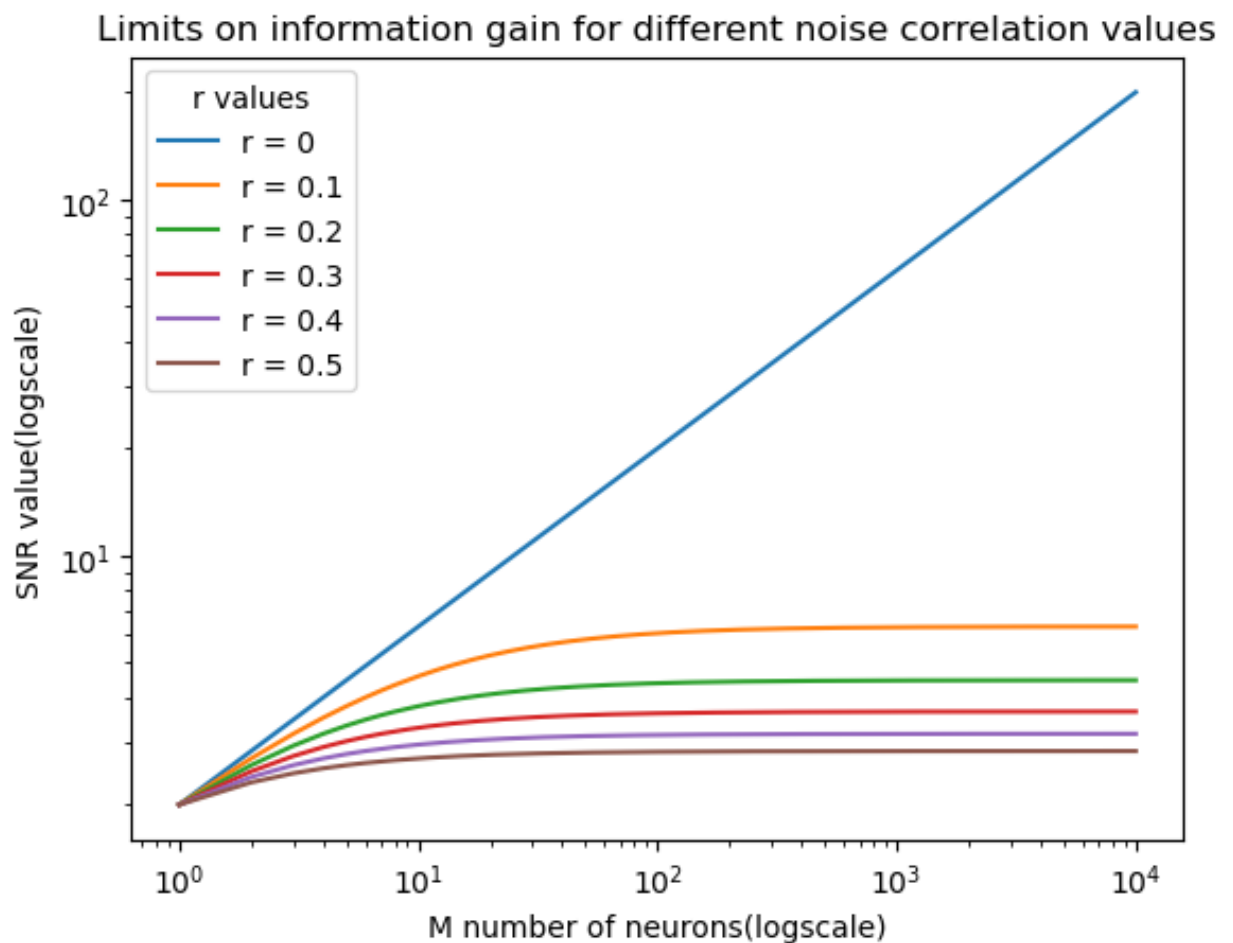
$$N = \sqrt{M \cdot \sigma^2 + M \cdot (M-1) \cdot r \cdot \sigma^2}$$

c) With your derivation of S and N, please calculate the signal to noise ratio S/N, and use this equation to replicate the plot on slide 8 of the second population code lecture, with x axis being the number of neurons (M) and y axis being the signal-to-noise ratio. Make sure to plot for a range of r values, going from $r = 0$ to $r = 0.5$. Also, be careful of the type of scale used for x and y axis.

```
In [2]: mean = 10
        stddev = 5
        M = np.linspace(1, 10000, 10000)
        S = M * mean
        for r in [0, 0.1, 0.2, 0.3, 0.4, 0.5]:
            N = np.sqrt(M * (stddev ** 2) + M * (M - 1) * r * (stddev ** 2))
            plt.plot(M, S/N)
        plt.xscale("log")
        plt.yscale("log")
        plt.legend(["r = 0", "r = 0.1", "r = 0.2", "r = 0.3", "r = 0.4", "r =
        plt.xlabel("M number of neurons(logscale)")
        plt.ylabel("SNR value(logscale)")
        plt.title("Limits on information gain for different noise correlation
        plt.show()
```


d) Prove either analytically or via simulation that when the number of neurons (M) is large enough, the best signal-to-noise ratio one can get is limited by $\frac{1}{\sqrt{r}}$.

$$N = \sqrt{M \cdot \sigma^2 + M \cdot (M-1) \cdot r \cdot \sigma^2}$$

$$N = \sqrt{\left((1-r) \cdot M + r \cdot M^2\right) \cdot \sigma^2}$$

$$\frac{S}{N} = \sqrt{\frac{M^2}{\left((1-r) \cdot M + r \cdot M^2\right)} \cdot \frac{\langle \mathbf{X} \rangle^2}{\sigma^2}}$$

$$\frac{S}{N} = \sqrt{\frac{1}{\left((1-r) \cdot \frac{1}{M} + r \cdot M\right)} \cdot \frac{\langle \mathbf{X} \rangle^2}{\sigma^2}}$$

For large M:

$$\frac{S}{N} = \sqrt{\frac{1}{r} \cdot \frac{\langle \mathbf{X} \rangle^2}{\sigma^2}}$$

This shows that the best signal to noise ratio which is represented by $\dfrac{\langle \mathbf{X} \rangle}{\sigma}$, gotten from the original equation for S/N by setting r to 0, is limited by the value $\dfrac{1}{\sqrt{r}}$ when $r > 0$.

In [ ]: