Homework 1
Systems & Theoretical Neuroscience [SWC]

Due: Wed, 18th October

1 Optimal sensory inference

1.1

Suppose you are a robot that has been teleported into a 100x100 metre arena. You want to estimate your position in the arena (x, y).

a) You initially do not know where in the arena you are, but from experience you believe it is slightly more likely that you are closer to the middle than to the walls of the arena. You choose to represent this prior belief as a Gaussian probability distribution. Draw a picture to show your prior belief. (you can represent a Gaussian by a centre point and a circle/ellipse showing the variance)

b) After waiting a few seconds, you finally get a GPS signal. The GPS tells you that you are located at position $(25m, 25m)$ in the arena. You know that GPS measurements are subject to noise with variance $10m^2$. Draw a picture to show the probability of the true position given the measurement.

c) You combine your prior belief and likelihood for the GPS measurement to update your belief of your position (this is called your posterior belief given the data). Draw a picture to show this new belief as another Gaussian. How has its mean and variance changed compared to your prior?

d) You lose GPS signal, but you have access to a distance sensor. It tells you that you're 70 metres from the northern wall, with a variance of $1m^2$. Draw a picture to show your updated posterior belief in your position. How has its mean and variance changed compared to your prior?

1.2

Perception can be seen as a problem of inference: given my sensory input and prior experiences, what is my 'best guess' of what is in the world? This idea can be tested in the estimation of motion velocity. In an experiment, human subjects were asked to complete a two-alternative forced choice (2AFC) task. They were presented with with moving grating stimuli (figure 1) and were tasked to judge which stimulus was moving faster.

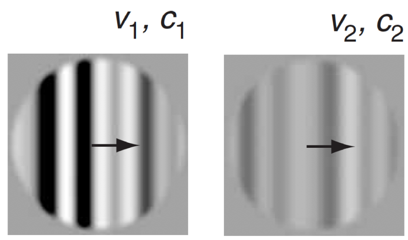

Figure 1: Experimental stimuli in the 2AFC task. Both gratings are moving in the same direction and subjects have to guess which stimulus moves faster. Both the velocity and the contrast are experimentally varied.

a) Performance on the 2AFC task showed that subjects systematically underestimated the velocity of low-contrast stimuli, while being more accurate when estimating the velocity of high-contrast stimuli. How would you explain this in terms of prior beliefs?

b) Write down Bayes' rule for inferring the posterior distribution from a likelihood function of measurements $\vec{m}$ given velocities v, and a prior belief over these velocities.

When humans solve the 2AFC task described above, they independently form an optimal estimate of the velocities of both stimuli, v_1 and v_2, yielding two posterior distributions $p(v_1|\vec{m_1})$ and $p(v_2|\vec{m_2})$. The mean of this posterior distribution reflects the observers best guess of the stimulus speed: $\hat{v}$. The posterior distribution and its mean $\hat{v}$ on every trial will vary due to measurement noise. We denote the distribution of estimates for a given stimulus speed as $p(\hat{v}|v)$.

c) The probability that stimulus 2 moved faster than stimulus 1 can be described with a cumulative probability function $p(\hat{v_2} > \hat{v_1})$. What does the cumulative distribution function for a Gaussian distribution look like? What happens to its shape when you increase and decrease the width of the Gaussian? To answer this question, have a look at the Jupyter notebook that was included in this assignment[1].

d) Psychometric curves such as the one in the Jupyter notebook are used to relate the physical properties of a stimulus to responses in a forced choice task. What effect do you think contrast variations will have on the psychometric curve?

[1]Instructions for installing everything you need to run the Jupyter notebook are attached on the next page. The Notebook can be found on `https://github.com/SWC-Gatsby-SNTN/assignment1`

1.3

a) Let $f(x)$ and $g(x)$ be two Gaussian probability density functions in x with means μ_1, μ_2 and variances σ_1^2, σ_2^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right); \qquad g(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$

Show that the product of the two PDFs is also a Gaussian $\mathcal{N}(\mu_3, \sigma_3)$, with mean and variance defined as:

$$\mu_3 = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}; \qquad \sigma_3^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

b) You are trying to estimate the width of a birthday present. Your prior belief is that birthday present widths (in cm) are distributed as a Gaussian $P(w) \sim \mathcal{N}(20, 10)$.

i) You query your visual system and it tells you the present is 15cm wide. You know that your visual system has an error variance of 5cm^2.

Write down the measurement likelihood $p(\hat{w}_V \mid w)$ where $\hat{w}_V$ is the estimate provided by vision.

ii) Use Bayes' rule (and the results derived above) to compute the posterior estimate $P(w \mid \hat{w}_V)$.

iii) You query your somatosensory system and it tells you the present is 18cm wide. You know that your somatosensory system has an error variance of 1cm^2.

Use Bayes' rule to compute the new posterior estimate $P(w \mid \hat{w}_S, \hat{w}_V)$.

iv) Your somatosensory system was much more accurate at estimating the width of an object. Did you gain anything by including your much noisier visual estimate?

Instructions for running the Jupyter notebook

If you have never used Python or Jupyter before, here is a quick guide to installing all the necessary requirements.

a) There are many ways to install Python. A quick and easy way is by installing Miniconda: `https://conda.io/miniconda.html`.

b) Second, you will need to install the Python package manager pip. This will provide you with an easy way to install packages from the Python package index (PyPI). With Conda installed, install pip by opening a terminal (Mac & Linux) or command prompt (Windows) and typing:

```
conda install pip
```

c) Next, use pip to install the packages required for the Jupyter notebook, and activate the ipywidgets:

```
pip install jupyter numpy matplotlib ipywidgets
jupyter nbextension enable --py widgetsnbextension
```

d) To run the Jupyter notebook, cd into the folder where the notebook is saved, and type:

```
jupyter notebook
```

This will open up your browser, and you can open up the notebook by clicking on it.

2 Point Processes

2.1

Figure 2 and Figure 3 show an example of a homogeneous and inhomogeneous rate function, respectively. The spike train (spike times are indicated by bars) in each figure was generated from a Poisson Process with the respective rate function.

a) Comment on the qualitative differences in the two rates and resulting spike trains. Why might a neuron fire at an inhomogeneous rate? Give examples.

b) Would you expect all spike trains that are generated from these rate functions to look the same? If yes, explain why. If no, explain where this variability comes from and what differences you might expect to see.

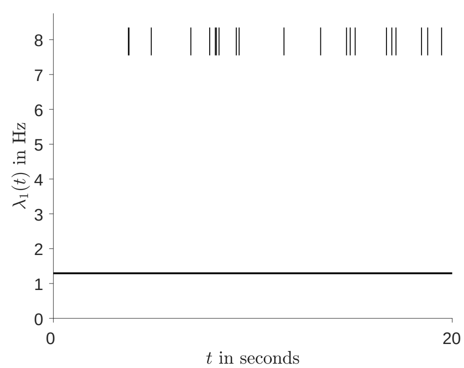

Figure 2: Homogeneous rate $\lambda_1(t)$ with spike train.

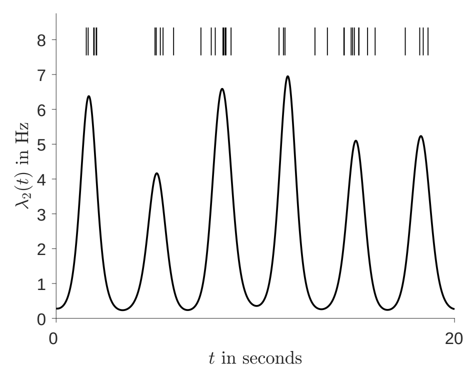

Figure 3: Inhomogeneous rate $\lambda_2(t)$ with spike train.

2.2

Suppose we would like to model the neurons in question 1.

a) What properties of neural firing would you like to capture in a model? Why does it matter?

b) Suppose you have decided to start with a model that assumes a constant rate. You measure the spike count in time bins $\Delta t = 200$ms in width. You call these spike counts $\{y_1, \ldots, y_N\}$ where $N = 100$. Your model is

$$y_i|\lambda \overset{iid}{\sim} \text{Poisson}(\lambda\Delta t), \qquad \text{for } i = 1, \ldots, N$$

(This notation means: each y_i is sampled independently from a Poisson distribution with the rate parameter $\lambda\Delta t$.)

Write down the log-likelihood of the spike counts as a function of the unknown rate parameter. Hint: The density of a Poisson distributed random variable $y|\theta \sim \text{Poisson}(\theta)$ is

$$p(y|\theta) = \frac{1}{y!}\theta^y \exp(-\theta)$$

c) Differentiate your likelihood with respect to λ, set to zero and solve to obtain a maximum likelihood estimate.

d) For the two spike trains in Figures 2 and 3 above, you have that $\sum_{i=1}^N y_i^{(1)} = 21$ for Figure 2 and $\sum_{i=1}^N y_i^{(2)} = 38$ for Figure 3. Plug this into your maximum likelihood estimator. Does the estimate agree with the true underlying rates?

e) How could you change your model to provide a better estimate for $\lambda_2(t)$? Write down a model and explain your choices.

f) Suppose you have derived an estimator for your new model and you now have two sets of potentially different estimates for λ_1 and λ_2. How can you decide which one is better?

3 Information Theory

3.1 Introduction

Information theory is a field that started with a groundbreaking paper by Shannon in 1948. This provided a formal understanding of what information is, and also outlined the basic constraints of transmitting information. This has significant implications in Neuroscience in terms of understanding the information content in neuronal responses, as well as the basic evolutionary constraints (e.g. the cost of information to a neuron or system of neurons). (For more information see Principles of Neural Information Theory: Computational Neuroscience and the Efficient Coding Hypothesis.. unfortunately there is only currently one chapter or so available online).

3.2 Intuition and basic examples

a) Before proceeding, read this detailed and intuition-led guide to information theory:
`http://colah.github.io/posts/2015-09-Visual-Information/`

b) Provide a qualitative explanation and (where applicable) a mathematical equation for the following:

- what is information and how can you measure it?
- what is entropy?
- what are the fundamental limits on storage and transmission of information? hint: what is channel capacity?
- what is mutual information?
- describe the relationship between mutual information and the Kullback-Leibler divergence.

c) Compute the entropy of a fair coin.

d) You have a set of coins, each with a different probability of coming up heads. Plot entropy as a function of probability.

e) You have an unfair die with the following probabilities:

P(X=1) = 1/9
P(X=2) = 1/6
P(X=3) = 1/6
P(X=4) = 0
P(X=5) = 2/9
P(X=6) = 1/3

Calculate the entropy of this system.

f) Let X and Y represent the responses of two neurons to a photograph of Tom Sellick's moustache (for simplicity each neuron can either be classed as a responder or not). Compute the entropy of the joint distribution $P(X, Y)$ given by:

P(N1 active, N2 active) = 1/2
P(N1 active, N2 inactive) = 1/4
P(N1 inactive, N2 active) = 1/4
P(N1 inactive, N2 inactive) = 0

Calculate the entropy of the joint distribution $P(X, Y)$

4 BONUS: Redundant and Synergistic Codes

Consider the problem of estimating the position S of your right arm from a pair of internal (e.g. proprioceptive, visual) signals R_1, R_2. As you (presumably) explained above, one measure of how informative this pair of signals is about S is the *mutual information* between S and R_1, R_2:

$$I[S; R_1, R_2] = \int P(S, R_1, R_2) \log \frac{P(S, R_1, R_2)}{P(S)P(R_1, R_2)}$$

a) Is it necessarily true that $I[S; R_1, R_2] = I[S; R_1] + I[S; R_2]$? Discuss intuitively why this might not be true.

b) Let's actually put some labels on R_1 and R_2 and verify our intuitions. Let R_1 be an efferent copy of the motor command signal that moved your right arm to its current location, and R_2 be proprioceptive input from muscle spindle afferents in your right biceps. We now note two statistical properties of these signals that are highly relevant to their coding capabilities:

   - In general, arm motor commands will be highly correlated with proprioceptive input from your bicep, since almost any arm movement involves lengthening or shortening of the biceps. We thus assume that R_1 and R_2 are *not* independent. In other words,

   $$P(R_1, R_2) = P(R_1)P(R_2|R_1) = P(R_2)P(R_1|R_2) \neq P(R_1)P(R_2)$$

   (where I have also sneaked in the probability product rule to give you a hint)

   - If we knew the arm's position S, knowing the efferent motor command R_1 would tell us nothing more about R_2. And vice versa, if we knew S, proprioceptive input would tell us nothing extra about the motor command R_1. We thus say that R_1 and R_2 are *conditionally independent* given S:

   $$P(R_1, R_2|S) = P(R_1|S)P(R_2|S)$$

   Use these two properties to show that $I[S; R_1, R_2] < I[S; R_1] + I[S; R_2]$ (recall that for any X, Y, $I[X; Y] \geq 0$).

   In this case, we say that the code provided by our pair of signals R_1, R_2 is *redundant*.

c) Consider now the case of R_1,R_2 being the responses from muscle spindle afferents in your right deltoid (a shoulder muscle) and biceps, respectively. We note the following properties:

   - Since almost any given arm position can be achieved with many different patterns of deltoid and biceps activity, R_1 and R_2 are roughly uncorrelated. We thus assume that R_1 and R_2 are *marginally independent*:

   $$P(R_1, R_2) = P(R_1)P(R_2)$$

   - If we knew the arm position S, knowing the deltoid activity R_1 would allow us to work out what the corresponding biceps activity R_2 was for the arm to be at point S. And vice versa, knowing the biceps activity R_2 would allow us to work out what the deltoid activity R_1 was. Thus, R_1 and R_2 are *not* conditionally independent:

   $$P(R_1, R_2|S) \neq P(R_1|S)P(R_2|S)$$

Use these two properties to show that $I[S; R_1, R_2] > I[S; R_1] + I[S; R_2]$.

In this case, we say that the code provided by our pair of signals R_1, R_2 is *synergistic*.

d) Now work out the same problem for the case where R_1 and R_2 are both marginally and conditionally independent:

$$P(R_1, R_2) = P(R_1)P(R_2)$$

$$P(R_1, R_2|S) = P(R_1|S)P(R_2|S)$$

E.g. R_1 = left biceps proprioceptive input, R_2 = right biceps proprioceptive input. How is $I[S; R_1, R_2]$ related to $I[S; R_1]$ and $I[S; R_2]$ in this case?