

Assignment 5  
Theoretical Neuroscience 2023  
Maneesh Sahani

TAs:

Kira Düsterwald (kira.dusterwald.21@ucl.ac.uk)  
Samo Hromadka (samo.hromadka.21@ucl.ac.uk)  
Dimitri Meunier (dimitri.meunier.21@ucl.ac.uk)  
Aaditya Singh (aaditya.singh.21@ucl.ac.uk)

**Due: 8 March 2023**

**1. Estimation Theory**

- (a) We derived the Fisher information  $J(\theta)$  as the expected value of the second derivative (curvature) of the log-likelihood in the lecture.
- i. Repeat the derivation for a *vector* parameter (or stimulus in our setting)  $\vec{\theta}$ , showing that the Fisher information in this case is given by a matrix.
  - ii. As mentioned in the lecture, there is an alternate definition in terms of the first derivative. For vector parameters this is:

$$J(\vec{\theta}_0) = \text{Cov}_{\vec{\theta}_0} \left( \nabla \log p(n|\vec{\theta}) \Big|_{\vec{\theta}_0} \right) \quad (1)$$

where  $\text{Cov}_{\vec{\theta}_0}$  means the covariance evaluated under  $p(n|\vec{\theta}_0)$ .

Demonstrate that these two definitions are the same (or more precisely, give conditions under which these two definitions are the same).

- (b) Consider an LNP model:

$$p(n|\vec{x}) = \text{Poisson}(g(\vec{w} \cdot \vec{x})) \quad (2)$$

- i. What is  $J(\vec{x})$  (the Fisher Information about the stimulus value available to the rest of the brain)? How does it depend on  $\vec{w}$ ? Working in two dimensions, show how  $J(\vec{x})$  varies around the vector linear projection vector  $\vec{w}$ .
- ii. What is  $J(\vec{w})$  (the Fisher Information about the weight vector available to an experimenter — consider the case of multiple measurements  $n_i$ , each in response to a different stimulus  $\vec{x}_i$ )? How does it depend on the distribution of  $\vec{x}$ ?
- iii. What would be a good distribution with which to probe the cell if we knew (say) the orthant of stimulus space in which  $\vec{w}$  lay?

## 2. Population Coding

Shadlen and collaborators have claimed that if the activities of neurons in population codes are corrupted by *correlated* noise, then there is a sharp limit to the useful number of neurons in the population. *Prima facie* this is wrong – the stronger the correlations, the lower the entropy of the noise, and therefore the stronger the signal.

Resolve this issue for the case of additive and multiplicative noise by considering the following three models for the noisy activities  $r_1$  and  $r_2$  of two neurons which form a population code for a real-valued quantity  $x$ :

$$\text{i.} \quad \begin{cases} r_1^a = x + \epsilon_1 \\ r_2^a = x + \epsilon_2 \end{cases} \quad (3)$$

$$\text{ii.} \quad \begin{cases} r_1^b = x(1 - \delta) + \epsilon_1 \\ r_2^b = x(1 + \delta) + \epsilon_2 \end{cases} \quad (4)$$

$$\text{iii.} \quad \begin{cases} r_1^c = x(1 - \delta)(1 + \eta_1) \\ r_2^c = x(1 + \delta)(1 + \eta_2) \end{cases} \quad (5)$$

where  $\delta \neq 0$  is known, and,  $\epsilon$  and  $\eta$  are Gaussian, with mean 0 and covariance matrices:

$$\Sigma = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$$

(a) What is the maximum likelihood estimator (MLE) for  $x$  on the basis of  $r_1$  and  $r_2$  in each case?

(b) What is the appropriate measure of the expected accuracy of the MLE, and why?

[*Hint: begin by showing that the Fisher Information for a Gaussian distribution with mean  $\mu(\theta)$  and variance  $\Sigma(\theta)$  both dependent on a scalar parameter  $\theta$  is:*

$$J(\theta) = \nabla\mu^\top \Sigma^{-1} \nabla\mu + \frac{1}{2} \text{Tr} [\Sigma^{-1} (\nabla\Sigma) \Sigma^{-1} (\nabla\Sigma)] \quad (6)$$

*where the matrix “gradient” is the matrix of element-wise derivatives. How does the Fisher Information help us to bound accuracy of an unbiased estimator?*

(c) How does the expected accuracy in each case i, ii, iii depend on the degree of correlation  $c$ ?

(d) What conclusions would you draw about the clash between Shadlen and common sense?