Week 1 - Neural code and Fisher information

The goal of the first week is to:

- Get familiar with the Fisher information and its application to neuronal tuning
- Compute the Fisher information for a single neuron and population of neurons with unimodal tuning curves
- Understand the scaling rules for neuronal tuning width

1 Welcome to NX-414: some notes on organization

A note on formatting

We will provide additional pointers and materials (also from the lectures) in the problem sets. To make clear which parts are actually exercises that you should work on, we put them as blue boxes...

2 Fisher information

When studying parameter estimation problems, we obtain information about the parameter from a sample of data coming from the underlying probability distribution. We wonder how much information can a sample of data provide about the unknown parameter. Fisher information is a measure of the amount of information that an observable random variable X carries about an unknown parameter θ.

We consider a random variable X for which the probability density function (pdf) or probability mass function (pmf) is $P(x|\theta)$, where θ is an unknown parameter and $\theta \in \Theta$, where Θ is the parameter space. We define $l(x|\theta) = \log P(x|\theta)$ as a log-likelihood function, therefore:

$$l'(x|\theta) = \frac{\partial}{\partial\theta} \log P(x|\theta) = \frac{P'(x|\theta)}{P(x|\theta)} \tag{1}$$

Intuitively, if an event has small probability, then the occurrence of this event brings us much information. For example, if P is sharply peaked with respect to changes in θ, it is easy to indicate the "correct" value of θ from the data, or equivalently, that the data X provides a lot of information about the parameter θ. If P is flat and spread-out, then it would take many samples of X to estimate the actual "true" value of θ that would be obtained using the entire population being sampled.

Below, we present three methods to calculate Fisher information.

Fisher information (for θ) contained in a random variable X is defined as:

$$I(\theta) = \mathbb{E}_\theta \left[(l'(X|\theta))^2\right] = \int \left(\frac{\partial}{\partial\theta} \log(P(x|\theta))\right)^2 P(x|\theta)dx \tag{2}$$

$$I(\theta) = Var_\theta[l'(X|\theta)] \tag{3}$$

$$I(\theta) = -\mathbb{E}_\theta[l''(X|\theta)] = -\int \frac{\partial^2}{\partial\theta^2} \log(P(x|\theta))P(x|\theta)dx \tag{4}$$

Exercise 1.1

Suppose a random variable X has a Bernoulli distribution for which the parameter θ is unknown ($0 < \theta < 1$). Find the Fisher information $I(\theta)$ in X.

Exercise 1.2

Suppose a random variable X has a Gaussian distribution for which μ is unknown, but the value of σ^2 is given. Find the Fisher information $I(\sigma)$ in X.

Exercise 1.3

Suppose a random variable X has a Poisson distribution for which the mean θ is unknown $(\theta > 0)$. Find the Fisher information $I(\theta)$ in X.

3 Fisher information of a Poisson neuron with a Gaussian tuning curve

Exercise 2.1

Compute the Fisher information for a Poisson neuron with Gaussian tuning curve in 1D. More details are given below. The number of spikes emitted by a certain neuron in a fixed interval approximately follows a Poisson probability distribution:

$$P(k|\mathbf{x}) = \frac{(\lambda(\mathbf{x}))^k}{k!} \exp\left(-\lambda(\mathbf{x})\right) \tag{5}$$

where $\lambda(\mathbf{x})$, indicates the average number of spikes emitted by the neuron, is a function of the encoded variable $\mathbf{x} \in \mathbb{R}^D$ and k is the number of spikes. For example, if the firing rate of the neuron depends on the location of the animal in space, $\mathbf{x}$ would include, among others, the Cartesian coordinates of the animal. For the first part of the exercise, we assume that x is a scalar value (i.e. $D = 1$) and that $\lambda(x)$ has the following expression:

$$\lambda(x) = f_M \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{6}$$

where μ is the value of x at which the spiking probability is maximum (sometimes called preferred stimulus), σ is a measure of how concentrated the firing probability is around μ, and f_M is the average number of spikes emitted when $x = \mu$. Overall, λ follows a rescaled Gaussian distribution.
Compute the expressions of $\frac{\partial \log(\lambda(x))}{\partial x}$ and $\frac{\partial \lambda(x)}{\partial x}$ as functions of x and of the parameters μ, σ and f_M

Exercise 2.2

Use the expressions computed at the previous point to find the Fisher information $I(x)$ of the number of spikes emitted by a single neuron, as a function of the variable x.[a]

[a]Hint: the variance of a Poisson distribution with parameter λ is λ. It might be convenient to isolate the expression of the variance of a Poisson random variable to easily compute the expectation in the definition of the Fisher information).

Exercise 2.3

We now want to compute the average Fisher information across multiple neurons. For simplicity, we assume that the activities of all neurons are independent from each other, that all neurons have identical tuning parameters and that their tuning centers μ are distributed with constant density η in the space of the encoded variables $\mathbb{R}^D$. Under these assumptions, the Fisher information can be computed with the following integral:

$$I = \eta \int_{-\infty}^{+\infty} I(x)dx \tag{7}$$

Compute the Fisher Information I as a function of the parameters η, f_M and σ.[a]

[a]Hint: remember that λ follows a rescaled Gaussian distribution. You might find it convenient to manipulate the integral as to isolate the expression of the variance of a Gaussian distribution of mean μ and variance σ^2.

Exercise 2.4

Let's now consider a multi-dimensional encoded variable space ($D > 1$). In this case, $\mathbf{x}, \mu \in \mathbb{R}^D$. We consider the following tuning curve for λ:

$$\lambda(\mathbf{x}) = f_M \exp\left(-\frac{||\mathbf{x}-\mu||^2}{2\sigma^2}\right) \tag{8}$$

where we have assumed that the average number of spikes emitted by the neuron depends on D independent encoded variables with the same variance. With the same assumptions as in the case $D = 1$, compute the Fisher information matrix for whole neural population, knowing that the element (i, j) of the Fisher information matrix as a function of the encoded variable $\mathbf{x}$ is defined as:

$$I_{ij}(\mathbf{x}) = \mathbb{E}\left[\frac{\partial \log(P(n|\mathbf{x}))}{\partial x_i}\frac{\partial \log(P(n|\mathbf{x}))}{\partial x_j}\right] \tag{9}$$

Does the result reflect the findings by Zhang and Sejnowksi (Neural Computation 1999) about the relation between tuning width and Fisher information, also reported in the lesson slides?