NX-414
A. Mathis
M. Schrimpf

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 $\mathcal{L}(\theta|x) = \log P(x|\theta)$ as a log-likelihood function, therefore:

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

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:

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

$$I(\theta) = Var_X\left[\frac{\partial}{\partial \theta}\mathcal{L}(\theta|X)\right] \tag{3}$$

$$I(\theta) = -\mathbb{E}_X\left[\frac{\partial^2}{\partial\theta^2}\mathcal{L}(\theta|X)\right] = -\int \frac{\partial^2}{\partial\theta^2}\log(P(x|\theta))P(x|\theta)dx. \tag{4}$$

We remark that the expected values and the variance listed in (2), (3) and (4) are computed with respect to X (in fact, $\boldsymbol{\theta}$ is a parameter and not a random variable). In the rest of the document, we indicate random variables with capital letters.

Exercise 1.1

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

Solution:

$$P(k|\theta) = \begin{cases} \theta, & \text{if } k = 1 \\ 1-\theta, & \text{if } k = 0 \end{cases} \quad \text{where } 0 < \theta < 1 \tag{5}$$

$$P(k|\theta) = \theta^k(1-\theta)^{(1-k)} \tag{6}$$

$$\log(P(k|\theta)) = k\log(\theta) + (1-k)\log(1-\theta) \tag{7}$$

$$\frac{\partial \log(P(k|\theta))}{\partial\theta} = \frac{k}{\theta} - \frac{1-k}{1-\theta} = \frac{k-\theta}{\theta(1-\theta)} \tag{8}$$

$$I(\theta) = \mathbb{E}\left[\frac{(K-\theta)^2}{\theta^2(1-\theta)^2}\right] = \frac{1}{\theta^2(1-\theta)^2}Var[K] = \frac{1}{\theta(1-\theta)} \tag{9}$$

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 $I(\mu)$ in X.

Solution:

$$P(x|\mu) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{10}$$

$$\log((P(x|\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2} \tag{11}$$

$$\frac{\partial \log(P(x|\mu))}{\partial\mu} = \frac{x-\mu}{\sigma^2} \tag{12}$$

$$I(\mu) = \mathbb{E}\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^4}Var[X] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} \tag{13}$$

Exercise 1.3

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

Solution:

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{14}$$

$$\log(P(k|\lambda)) = k\log(\lambda) - \lambda - \log(k!) \tag{15}$$

$$\frac{\partial \log(P(k|\lambda))}{\partial \lambda} = \frac{k}{\lambda} - 1 \tag{16}$$

$$I(\lambda) = \mathbb{E}\left[\left(\frac{\partial \log(P(K|\lambda))}{\partial \lambda}\right)^2\right] = \mathbb{E}\left[\left(\frac{K-\lambda}{\lambda}\right)^2\right] = \frac{1}{\lambda^2}Var[K] = \frac{1}{\lambda} \tag{17}$$

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 K emitted by a certain neuron in a fixed interval is a random variable approximately following a Poisson probability distribution:

$$P_K(k|\boldsymbol{\theta}) = \frac{(\lambda(\boldsymbol{\theta}))^k}{k!} \exp\left(-\lambda(\boldsymbol{\theta})\right) \tag{18}$$

where $\lambda(\boldsymbol{\theta})$ indicates the average number of spikes emitted by the neuron and is a function of the encoded variable $\boldsymbol{\theta} \in \mathbb{R}^D$[a]. For example, if the firing rate of the neuron depends on the location of the animal in space, $\boldsymbol{\theta}$ 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 $\lambda(\theta)$ has the following expression:

$$\lambda(\theta) = f_M \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right) \tag{19}$$

where μ is the value of θ 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 $\theta = \mu$. Overall, λ follows a rescaled Gaussian distribution.
Compute the expressions of $\frac{\partial \log(\lambda(\theta))}{\partial \theta}$ and $\frac{\partial \lambda(\theta)}{\partial \theta}$ as functions of θ and of the parameters μ, σ and f_M

[a] In neuroscience publications the variable $\boldsymbol{\theta}$ is often indicated as $\mathbf{x}$. Here we use $\boldsymbol{\theta}$ for consistency with our previous notation and with the statistics literature.

Solution:

$$\log(\lambda(\theta)) = -\frac{(\theta-\mu)^2}{2\sigma^2} + C \tag{20}$$

where C is constant with respect to θ.

$$\frac{\partial \log(\lambda(\theta))}{\partial\theta} = -\frac{\theta-\mu}{\sigma^2} \tag{21}$$

$$\frac{\partial\lambda(\theta)}{\partial\theta} = f_M \exp\left(\frac{(\theta-\mu)^2}{2\sigma^2}\right)\left(-\frac{\theta-\mu}{\sigma^2}\right) = -\lambda(\theta)\frac{\theta-\mu}{\sigma^2} \tag{22}$$

Exercise 2.2

Use the expressions computed at the previous point to find the Fisher information $I(\theta)$ in the number of spikes K emitted by a single neuron. Report the result as a function of the variable θ.[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).

Solution: We want to compute the information that the observable random variable representing the number of spikes emitted by the neuron (in this solution we call such random variable K) carries about the unknown parameter θ. K is distributed as a Poisson of parameter $\lambda(\theta)$, therefore it depends on θ through its parameter λ. For this purpose, we use the definition of Fisher information and the derivatives computed at the previous point:

$$\begin{aligned} I(\theta) = \mathbb{E}\left[\left(\frac{\partial \log P(K|\theta)}{\partial\theta}\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\left(K\log(\lambda(\theta)) - \lambda(\theta) - \log(K!)\right)\right)^2\right] = \\ \mathbb{E}\left[\left(-K\frac{\theta-\mu}{\sigma^2} + \lambda(\theta)\frac{\theta-\mu}{\sigma^2}\right)^2\right] = \frac{(\theta-\mu)^2}{\sigma^4}\,\mathbb{E}\left[(K-\lambda(\theta))^2\right] = \frac{(\theta-\mu)^2}{\sigma^4}\lambda(\theta) \end{aligned} \tag{23}$$

where in the last step we have used the fact that K is a Poisson of parameter $\lambda(\theta)$, whose variance is in fact $\lambda(\theta)$.

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:

$$I = \eta\int_{-\infty}^{+\infty} I_\mu(\theta)d\mu \tag{24}$$

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.

Solution:

$$I(\theta) = \eta \int_{-\infty}^{+\infty} I_\mu(\theta) d\mu = \eta \int_{-\infty}^{+\infty} \frac{(\theta-\mu)^2}{\sigma^4} \lambda(\theta) d\mu = \eta \int_{-\infty}^{+\infty} \frac{(\theta-\mu)^2}{\sigma^4} f_M \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right) d\mu = \frac{\eta f_M \sqrt{2\pi}}{\sigma^3} \int_{-\infty}^{+\infty} (\theta-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right) d\mu = \frac{\eta f_M \sqrt{2\pi}}{\sigma} \tag{25}$$

In the last step we used the fact that the integral is the expression of the variance of a Gaussian distribution with mean θ and variance σ^2.

Exercise 2.4

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

$$\lambda(\boldsymbol{\theta}) = f_M \exp\left(-\frac{||\boldsymbol{\theta}-\boldsymbol{\mu}||^2}{2\sigma^2}\right) \tag{26}$$

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 $\boldsymbol{\theta}$ is defined as:

$$I_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \log(P(K|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log(P(K|\boldsymbol{\theta})}{\partial \theta_j}\right] \tag{27}$$

where K is the random variable associated to the number of spikes emitted by a single neuron.
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?

Solution: Proceeding similarly to the case $D = 1$, we have

$$\begin{aligned} I_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \log(P(K|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log(P(K|\boldsymbol{\theta})}{\partial \theta_j}\right] &= \mathbb{E}\left[(\lambda(\boldsymbol{\theta}) - K)^2 \left(\frac{\theta_i - \mu_i}{\sigma^2}\right)\left(\frac{\theta_j - \mu_j}{\sigma^2}\right)\right] \\ &= \lambda(\boldsymbol{\theta}) \left(\frac{\theta_i - \mu_i}{\sigma^2}\right)\left(\frac{\theta_j - \mu_j}{\sigma^2}\right) \end{aligned} \tag{28}$$

we now need to integrate such function of $\boldsymbol{\theta}$ over the whole space $\mathbb{R}^D$, again, under the assumption of constant density η. We perform the following partial computation:

$$\int_{\mathbb{R}} \exp\left(-\frac{(\theta_i-\mu_i)^2}{2\sigma^2}\right) = \sigma\sqrt{2\pi} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta_i-\mu_i)^2}{2\sigma^2}\right) = \sigma\sqrt{2\pi} \tag{29}$$

where we used the fact that the integral of a density function over its support is 1. We then observe that:

$$\int_{\mathbb{R}} (\theta_i - \mu_i) \exp\left(-\frac{(\theta_i-\mu_i)^2}{2\sigma^2}\right) = \sigma\sqrt{2\pi} \int_{\mathbb{R}} (\theta_i - \mu_i) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta_i-\mu_i)^2}{2\sigma^2}\right) = 0 \tag{30}$$

Here we used the fact that the integral computes the mean of a Gaussian distribution with mean equal to 0. Finally:

$$\int_{\mathbb{R}} (\theta_i - \mu_i)^2 \exp\left(-\frac{(\theta_i - \mu_i)^2}{2\sigma^2}\right) = \sigma\sqrt{2\pi} \int_{\mathbb{R}} (\theta_i - \mu_i)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta_i - \mu_i)^2}{2\sigma^2}\right) = \sigma^3\sqrt{2\pi} \quad (31)$$

where we used the fact that the integral is the variance of a Gaussian distribution with variance σ^2. Therefore:

$$\begin{aligned} I_{ij} &= \eta \int_{\mathbb{R}^D} I_{ij}(\boldsymbol{\theta})\, d\mu_1 \cdots d\mu_D = \eta \int_{\mathbb{R}^D} \lambda(\boldsymbol{\theta}) \left(\frac{\theta_i - \mu_i}{\sigma^2}\right)\left(\frac{\theta_j - \mu_j}{\sigma^2}\right) d\mu_1 \cdots d\mu_D \\ &= \frac{\eta f_M}{\sigma^4} \int_{\mathbb{R}^D} (\theta_i - \mu_i)(\theta_j - \mu_j) \exp\left(-\frac{||\boldsymbol{\theta} - \boldsymbol{\mu}||^2}{2\sigma^2}\right) d\mu_1 \cdots d\mu_D \\ &= \frac{\eta f_M}{\sigma^4} \int_{\mathbb{R}^D} (\theta_i - \mu_i)(\theta_j - \mu_j) \prod_{k=1}^{D} \exp\left(-\frac{(x_k - \mu_k)^2}{2\sigma^2}\right) d\mu_1 \cdots d\mu_D \end{aligned} \quad (32)$$

In the last step we simply manipulated the exponential function, using the definition of norm. We can observe that the integrand is the product of functions depending on one of the integration variables each, so it can be solved as the product of D integrals in one dimension. When $i \neq j$, then the product is 0, as the integral along both θ_i and θ_j is 0 (Eq. 30). When $i = j$, then the integral in all directions but μ_i is given by Eq. 29, while in direction μ_i it is given by Eq. 31. Therefore, we have:

$$I_{ii} = \frac{\eta f_M}{\sigma^4}(\sigma\sqrt{2\pi})^{D-1}\sigma^3\sqrt{2\pi} = \eta f_M (2\pi)^{D/2}\sigma^{D-2} \quad (33)$$

independently of the index i. The Fisher information matrix is thus $I = \eta f_M (2\pi)^{D/2}\sigma^{D-2}\mathbb{I}$, where $\mathbb{I}$ is the identity matrix.