TD 10 – Neuronal Coding & Information Theory

Practical Information

TD Assistant
eponiatowski@clipper.ens.psl.eu

TD Material
https://github.com/esther-poniatowski/2223_UlmM2_ThNeuro

Goals of the TD

This TD aims to tackles some topics of coding in neuroscience :

1. *Mutual information*
2. *Fisher information*
3. *Bayesian inference*

1 Mutual Information

1.1 Characterizing the distribution of a discrete random variable

Let X be a discrete random variable, which can take values $\{x_1, ..., x_n\}$ with probability $p(x_i)$.

The **Shannon Entropy** of the distribution is defined as :

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i))$$

This is also called the ***information*** of the distribution (in the sense of Shannon), as it is related to the number of 'yes-no' questions which can allow to infer the outcome of one random experiment.

In the following questions, the values x_i correspond to the possible colors of a ball picked randomly from an urn. One observer picks one ball in the urn, and another player has to ask 'yes-no' questions to determine its color.

(1) Calculate the entropy and the average number of questions needed to find out the color of the ball in the following cases :
- Only one possible color is in the urn.
- Two possible colors are in the urn.
- Half the balls are red, one fourth are green and one fourth are blue.

Comment.

1.2 Mutual information between two discrete random variables

The brain can be viewed as a processing system gathering information about the environment through its sensors. One way to model this information processing is to consider that the events s (stimuli) in the environment are stochastic and that the brain activity r (neuronal response) is correlated with those external events.

The **Mutual Information** between stimulus and response quantifies how much observing one is informative about the other :

$$I(s,r) = H(s) - H(s|r) = H(r) - H(r|s) = H(r) + H(s) - H(s,r) = \sum_{s,r} p(s,r) \log\left[\frac{p(s,r)}{p(s)p(r)}\right]$$

In the following questions, give the mutual information between stimuli s and the neural activity r. Try to answer without calculation (give the answer in bits).

② When stimulus and response are uncorrelated.

③ When the stimulus is binary ($s = A, B$) with equal probability, and a single neuron reacts with binary activity $r = 0$ if $s = A$ and $r = 1$ if $s = B$.

When the neuronal code involves the joint activity of more than one neuron, it might contain some redundancy.

For a two-neurons code with two neurons, with activities $r_i(s),\ i = 1, 2$, **redundancy** can be quantified by the following metric :

$$R \equiv I(s, r_1) + I(s, r_2) - I(s, \{r_1, r_2\})$$

Consider the following case :

- For $s = A$: $r_1(A) = r_2(A) = 1$.
- For $s = B$: $r_1(B) = r_2(B) = 0$.

④ What is the mutual information $I(s, \{r_1, r_2\})$? Between the stimulus and only the first neuron, $I(s, r_1)$?

⑤ What is the redundancy in this particular case ?

1.3 Mutual information for continuous random variables

Entropy and mutual information can be extended to the case of continuous random variables (in which the sum becomes an integral), according to :

$$H(s) = \int \mathrm{d}s\, p(s) \log\left[p(s)\right] \quad (1)$$

$$I(s,r) = \int \mathrm{d}s\, \mathrm{d}r\, p(s,r) \log\left[\frac{p(s,r)}{p(s)p(r)}\right] \quad (2)$$

⑥ Compute the entropy of the Gaussian distribution :

$$P(r) = \frac{1}{(2\pi\sigma^2)^{1/2}}\, e^{-\frac{(r-r_0)^2}{2\sigma^2}}$$

Consider the following linear system :

$$r = ws + z$$

with :

- w a (positive) weight,
- s a scalar stimulus following a Gaussian distribution :

$$\rho(s) = \frac{1}{(2\pi c^2)^{1/2}}\, e^{-\frac{s^2}{2c^2}}$$

- z a Gaussian noise with zero mean and variance σ^2 :

$$P(z) = \frac{1}{(2\pi \sigma^2)^{1/2}}\, e^{-\frac{z^2}{2\sigma^2}}$$

7 Compute the mutual information $I(r, s)$ between the stimulus s and the neural response r.

Consider the model :

$$r = \sum_{j=1}^{N} w_j s_j + z$$

with Gaussian inputs of zero mean, $\langle s_j \rangle = 0$ for every j, and covariance matrix C :

$$\langle s_j s_{j'} \rangle = C_{j,j'}$$

8 Compute the mutual information $I(r, s)$ for this model.

Discuss the maximization of mutual information with respect to the choice of the weights $\{w_j, j = 1, ..., N\}$.

2 Fisher Information

Shannon's information is a measure of how much information the response of a neuron provides about the *whole stimulus space*.

However, individual neurons in the brain appear to be "tuned" to certain regions of the stimulus space.

It can therefore be useful to introduce a *local measure of information*.

The notion of ***precision*** of the information provided by a neuron about a stimulus s_0 aims to reflect how clear it is, given the response of that neuron, to tell discriminate this stimulus from similar ones.

This notion is characterized by the **Fisher information**, which can be defined in two equivalent ways.

2.1 Distance between probability distributions

A first metric of precision is related to the 'difference' between the responses of the neuron when the stimulus s_0 is presented compared to a nearby stimulus $s_0 + \delta s$, in terms of probability distributions. The more the distributions $p(r|s_0)$ and $p(r|s_0 + \delta s)$ overlap, the less 'precise' this information is.

The **Kullback-Leibler divergence** quantifies the distance between these two probability distributions :

$$KL(p||q) = \int \mathrm{d}x\, p(x) \log\left[\frac{p(x)}{q(x)}\right] \quad (3)$$

$$KL(p||q) \geq 0 \quad (4)$$

$$KL(p||q) = 0 \Leftrightarrow p = q \quad (5)$$

Note that the Kulback-Liebler divergence is a distance metric which has already been used to define the mutual information between s and r as the distance between $p(s, r)$ and $p(s)p(r)$.

In this framework, the **Fisher information** locally around the stimulus s_0 is defined as follows :

$$F(s_0) = \frac{d^2 KL(p(r|s_0)||p(r|s))}{ds^2}(s_0) = -\int \mathrm{d}r\, p(r|s_0) \frac{\partial}{\partial S} \log(p(r|s))(s_0) \tag{6}$$

9 Sketch the Kullback-Leibler divergence $KL(p(r|s_0)||p(r|s))$ as a function of s.

10 Explain why the Fisher metric (6) gives a measure of the information that r locally provides about the stimulus.

2.2 Variance of the locally optimal estimator

Another approach to quantify the local precision of the information builds upon an **estimator** $\widehat{s}(r)$ of the stimulus given the neuronal response.

Among all estimators, *locally unbiased estimators* are accurate on average for values of the stimulus close to s_0 :

$$\frac{\partial}{\partial S}\langle \widehat{s} \rangle(s_0) = 1 \qquad \text{with} \quad \langle \widehat{s} \rangle(s) = \int \widehat{s}(r) p(r|s)\, \mathrm{d}r$$

Such an estimator, although accurate on average, will generally not provide an exact estimate of the stimulus on each trial. A way to measure the 'precision' of such an estimator is through the *inverse of its **variance*** at s_0 :

$$\int \mathrm{d}r\, p(r|s_0)(\widehat{s}(r) - s_0)^2$$

The **Cramér–Rao bound** expresses a lower bound on the variance of unbiased estimators, which is at least as high as the inverse of the Fisher information. Equivalently, it expresses an upper bound on the precision (the inverse of variance) of unbiased estimators, which is at most the Fisher information.

Cauchy-Schwarz inequality :

$$\int f(x) g(x)\, \mathrm{d}x \leq \int f^2(x)\, \mathrm{d}x \int g^2(x)\, \mathrm{d}x$$

with equality if and only if $f(x) = a g(x)$

11 Using the fact that the estimator is unbiased and the Cauchy-Schwarz inequality, show that :

$$\int \mathrm{d}r\, p(r|s_0)(\widehat{s}(r) - s_0)^2 \int \mathrm{d}r\, p(r|s_0) \left(\frac{\partial}{\partial S} \log(p(r|s))(s_0) \right)^2 \geq 1$$

12 Using the fact that the probability distribution is normalized, justify that

$$F(s) = \int \mathrm{d}r\, p(r|s) \left[\frac{\partial}{\partial S} \log p(r|s) \right]^2 = -\int \mathrm{d}r\, p(r|s) \frac{\partial^2}{\partial S^2} \log p(r|s)$$

13 Using the case of equality in the Cauchy-Schwarz inequality, find a locally unbiased estimator whose variance is equal to the inverse of the Fisher Information.

2.3 Examples of Fisher local information for different response models

Models of a single neuron's response to a stimulus s can be assessed through the mean response $f(s)$ and the variance of the neuron's response $\sigma(s)^2$.

14 Qualitatively, how do you expect the Fisher Information to depend on $f(s)$ and $\sigma(s)$?

In the following cases, give the Fisher Information and determine the optimal estimator.

15 Neuron with a Poisson firing rate :

$$P(r|s) = \frac{f(s)^r}{r!}\, e^{-f(s)}$$

16 Neuron with Gaussian noise :

$$P(r|s) = \frac{1}{\sqrt{2\pi}\sigma(s)} e^{-\frac{(r-f(s))^2}{2\sigma(s)^2}}$$

17 Two independent neurons, with constant individual variance $\sigma(s) = 0$

3 Bayesian inference

Neuronal coding can be distributed in a population of neurons with distinct tuning curves.

Consider a population of N neurons with various tuning curves $f_i(s)$. When a stimulus s is presented, the neuron i emits spikes with a Poisson process of mean $f_i(s)$. The variability in the number of spikes generated during any time bin is independent across neurons.

Bayes' rule aims to provide the probability distribution of the stimulus from the observed pattern of neuronal responses across the population :

$$\mathbb{P}(s|\{n_i\}) = \frac{\mathbb{P}(\{n_i\}|s)\mathbb{P}(s)}{\mathbb{P}(\{n_i\})}$$

18 When a stimulus s is presented, what is the probability $\mathbb{P}(\{n_i\}|s)$ of observing a given pattern of spikes $\{n_i\}$?

In the following questions, the following assumptions hold :

- The prior on the stimulus $\mathbb{P}(s)$ is uniform.
- The tuning curves are all similar bell-shaped curves, with the preferred stimuli s_i of the various neurons evenly distributed across the stimulus range.

19 When a pattern of spikes $\{n_i\}$ was observed, propose an *estimate* of the stimulus that was presented.

20 How does the accuracy of this estimate depend on the height f_0 and width σ of the tuning curves and on the number of neurons ?

21 When a pattern of spikes $\{n_i\}$ was observed, what is the *probability distribution* over the presented stimulus ? Include all the terms which don't depend on s in a function $\Phi(\{n_i\})$.

In the following questions, the tuning curves are Gaussian :

$$f_i(s) = f_0\, e^{-\frac{(s-s_i)^2}{2\sigma^2}}$$

22 Show that $\mathbb{P}(s|\{n_i\})$ is also a Gaussian. What is its mean and variance ? How does the variance depend on the various parameters ? Under what conditions may the variance become infinitely small ?

Instead of responding to s, neurons respond to a jittered version of s, namely $\widehat{s}$, where $p(s|\widehat{s})$ is a Gaussian of variance σ_j^2. The number of spikes they fire is drawn from a Poisson distribution of mean $f_i(\widehat{s})$.

23 For a given stimulus s, is the variability still independent across neurons ? Does the variance of $p(s|\{n_i\})$ still become infinitely small in the conditions considered previously ?