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 useful in neuroscience.

Part 1 Mutual information

Part 2 Fisher information

Part 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 in $\{x_1, ..., x_n\}$, each 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).

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 Compute the entropy and the average number of questions needed to determine the color of the ball in the following cases :

- Only one color is in the urn.
- Two colors are in the urn.
- Half the balls are red, one fourth are green and one fourth are blue.

② Comment on the relation between the entropy and the number of 'yes-no' questions which can allow to infer the outcome of one random experiment.
Interpret the entropy as an expectancy.

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 :

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

For each case below, give the mutual information between stimuli s and the neural activity r, if possible without calculation (answer in bits).

③ When stimulus and response are uncorrelated.

④ When the stimulus is binary with equal probability, and a single neuron reacts to it deterministically :

$$s \in \{A, B\}, \quad \begin{cases} s = A \iff r = 0 \\ s = B \iff r = 1 \end{cases}$$

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

For a code with two neurons, with activities $r_i(s),\ i \in \{1,2\}$, **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 : $s \in \{A, B\}, \quad \begin{cases} s = A \iff r_1(A) = r_2(A) = 1 \\ s = B \iff r_1(B) = r_2(B) = 0 \end{cases}$

⑤ What is the mutual information $I(s, \{r_1, r_2\})$ between the stimulus and the joint responses of both neurons ? What is the mutual information $I(s, r_1)$ between the stimulus and only the first neuron ?

⑥ What is the redundancy in this particular case ?

1.3 Mutual information for continuous random variables

Entropy and mutual information can be extended to continuous random variables.
In this case, the sum becomes an integral over the range of values taken by the random variable :

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

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

⑦ 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 a neuron model whose response is linearly related to the stimulus :

$$r = ws + z$$

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 :

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

8 Compute the mutual information $I(r, s_1, ..., s_N)$ between the stimuli and the neural response.

Consider a neuron model whose response is related to a combination of stimuli :

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

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

$$\langle s_i s_j \rangle = C_{i,j}$$

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

10 Discuss the maximization of mutual information with respect to the choice of the weights $w_j,\ j \in \{1, ..., N\}$, under the constraint $\sum_{j=1}^{N} w_j^2 = 1$.

2 Fisher Information

Shannon's information is a *global measure of information* : it quantifies how the response of a neuron provides information about the *whole stimulus space*, i.e. the whole distribution of stimuli values.

However, individual neurons in the brain appear to be "tuned" to certain "portions" of the stimulus space. Therefore, it is useful to introduce a *local measure of information*.

Fisher information aims to characterize the **precision** of the information provided by a neuron about a *particular stimulus* s_0. Intuitively, this measure reflects how clearly the response of the neuron allows to discriminate this particular stimulus from "similar" stimuli (nearby values).

Fisher information can be defined in two equivalent ways, introduced below.

2.1 Distance between probability distributions

A first metric of precision is related to the "difference" (in terms of probability distributions) between the responses of the neuron when the stimulus s_0 is presented compared to a nearby stimulus $s_0 + \delta s$. 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** is commonly used to quantify the distance between two probability distributions over the same space of values ($x \in \mathcal{X}$) :

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

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

$$F(s_0) = \left.\frac{d^2 KL(p(r|s_0)||p(r|s))}{ds^2}\right|_{s_0} = -\int \mathrm{d}r\, p(r|s_0)\, \left.\frac{\partial}{\partial s} \log(p(r|s))\right|_{s_0} \tag{4}$$

(11) Prove that for any pair of distibutions p and q : $KL(p||q) \geq 0$.
Justify that the first derivative of $KL(p(r|s_0)||p(r|s))$ with respects to s cancels at s_0.

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

(13) Prove the equality (4).

(14) From the interpretation of the derivative, explain why Fisher metric 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 r.

An estimator $\widehat{s}(r)$ of s_0 is said to be *locally unbiased* if it is accurate on average for values of the stimulus close to s_0 :

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

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.

Reminder : 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 $f(x) = ag(x)$

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

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

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

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

17 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 $\sigma(s)^2$.

18 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.

19 Neuron with a Poisson firing rate :

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

20 Neuron with Gaussian noise :

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

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

3 Bayesian inference

Neuronal coding can be distributed over a population of N neurons with distinct tuning curves $f_i(s)$.

When a stimulus s is presented, each 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 inferred from the observed pattern of neuronal responses across the population :

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

(22) 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 questions below, the following assumptions hold :

- The prior on the stimulus $\mathbb{P}(s)$ is uniform.
- Each neuron i has a bell-shaped tuning curve, centered on its preferred stimulus s_i. Across the population, the bell shapes evenly cover the stimulus range.

(23) Propose an *estimate* of the stimulus which can be inferred when a pattern of spikes $\{n_i\}$ is observed, using the intuition that each neuron 'votes' for its preferred stimulus.

(24) 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 ? Use question (19).

(25) Determine the *probability distribution* over the stimulus when a given pattern of spikes $\{n_i\}$ is observed. Include all the terms which do not depend on s in a function $\Phi(\{n_i\})$.

In the following question, the tuning curves are Gaussian :

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

(26) 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 $f_i(\widehat{s})$.

(27) 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 ?