Problem set 11: Maximum entropy

BIO-369

Prof. Anne-Florence Bitbol
EPFL

1 Some examples

The principle of maximum entropy yields an inference method that allows us to construct models (in the form of probability distributions) consistent with certain observations (means of observables). The idea is to impose the least possible structure or bias in the model, apart from the consistency with these observations. Formally, this corresponds to maximizing the entropy under constraints imposed by the observations. This is very useful when one has a partial knowledge of the studied system, i.e. only some data are known.

a) Consider a discrete random variable X that can take all non-negative integer values. Assume that we have measured its mean m experimentally. What is the maximum-entropy probability distribution consistent with this measurement? Express $P(x)$ as a function of m and x only.

b) For a continuous random variable X, the results we obtained for discrete ones can be generalized, and the maximum-entropy probability density $p(x)$ takes the same form as $P(x)$ in the discrete case. Given this, write down the form of the maximum-entropy probability density $p(x)$ for a continuous random variable with measured mean m. Assume that x can take all positive real values. Imposing the normalization and mean constraints, express it as a function of m and x only. What is the name of the distribution obtained?

c) Still for a continuous random variable X, assume that you have measured not only $\langle x \rangle = m$ but also $\langle x^2 \rangle = q$. Also assume that x can take all real values. What is the form of the maximum-entropy probability density $p(x)$? What is its variance?

2 Maximum entropy modeling of neuroscience data

In Ref. [1], the activity of 40 neurons from the salamander retina was analyzed while visualizing repetitions of a movie, in the spirit of the experiments discussed in previous lectures. The movie lasting $T = 26.5$ s was repeated $N_r = 120$ times during the recording.

a) In Python, load the data in the text file `Data11.dat`. It contains a list of numbers (in a single column) in the following order: for each neuron, first the number 4000 is written, then an integer labeling the neuron (we will not use these two rows in our analysis), and then, the successive spiking times of the neuron considered are given in seconds. These times can range from 0 to 3180 s, which is the total time of the recording, namely $N_r \times T$. Then for the next neuron we have the same data (two useless rows, and then successive spiking times).

   Given this, produce an array containing the row indices where the data shifts from one neuron to the next one. You can then use it to extract the data for each neuron.

b) Extract the spiking times of the first neuron in the data. As the experiment comprises N_r repeats of movies lasting time T, use the `modulo` and `floor` operations on this data to obtain respectively:

   - an array `x` containing the times of firing with respect to the beginning of the movie repetition (i.e. with time origin being the beginning of the particular movie repetition), and
   - an array `y` containing the indices of the repetitions of the movie corresponding to each data point in `x`.

Make a scatter plot of `y` versus `x` and comment. You can look at the whole plot and then zoom to see individual markers corresponding to spikes.

c) Do the same thing for the second neuron in the dataset and compare the activity of these two neurons. Also bin the data with bins of duration 10 ms and plot a histogram of the number of bins where this second neuron is active versus time after the beginning of the movie.

d) For each neuron in the dataset, bin the data with bins of duration 10 ms. For each bin in the movie, obtain the frequency that the neuron spikes in this bin (recall that there are N_r repeats of the movie), ignoring the possibility that more than two spikes may occur in the same time bin. Store the data in an array `freqs` where each row is a time bin, and each column is a different neuron.

e) Does it happen that two spikes may occur in the same time bin? If yes, replace by one the frequencies that are larger than one.

f) We will now make the following mapping: a value 1 will be associated to a spiking neuron and a value -1 to an inactive neuron. Hence, the activity of neuron i is represented by a random variable X_i that can take values in $\{-1, 1\}$. Express the mean of X_i as a function of the probability that the neuron is active. Use this to produce an array containing the mean values of the activity of each neuron in each time bin.

g) Compute the mean activity over time for each neuron, averaged over bins b from 1 to B (the total number of bins) as

$$\langle x_i \rangle = \frac{1}{B} \sum_{b=1}^{B} x_i(b) \,. \tag{1}$$

Comment on the values, and make a scatter plot of these mean activities.

h) Compute the covariance of neuron activities for all pairs of neurons i and j, defined as:

$$C_{ij} = \mathrm{cov}(X_i, X_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \frac{1}{B} \sum_{b=1}^{B} x_i(b) x_j(b) - \langle x_i \rangle \langle x_j \rangle \,. \tag{2}$$

Plot the matrix of covariances in color, with the diagonal and then setting the diagonal elements to zero, and comment.

Also compute the matrix of correlation coefficients, and compute the mean correlation coefficient between two different neurons. Comment on the value obtained: do correlations appear important in this system?

i) The maximum entropy distribution consistent with the measurements of the mean activity of each neuron and of their covariances reads

$$P(x_1, x_2, \ldots, x_N) = \frac{1}{Z} \exp\left[-\sum_{i=1}^{N} h_i x_i - \sum_{i=1}^{N} \sum_{j<i} J_{ij} x_i x_j \right] , \tag{3}$$

where Z is a normalization constant. Comment on this formula.

Generally, inferring the values of the parameters h_i and J_{ij} consistent with the measurements is a difficult task. Thus, approximations have been developed. In one of them, known as the "small-coupling approximation" or the "mean-field approximation", the parameters J_{ij} are simply minus the elements of the inverse covariance matrix:

$$J_{ij} = C^{-1}_{ij} \,. \tag{4}$$

Estimate the parameters J_{ij} in this approximation. Plot their matrix in color, setting the diagonal elements to zero, and comment.

j) In Ref. [1], the parameters were estimated more precisely for groups of 10 neurons, and predictions from the maximum entropy distribution consistent with means and covariances was compared to those from a maximum entropy distribution that is just consistent with means. What would be the form of the probability distribution in such a model?

In particular, predictions were made from these two models for the rate at which each binary word or pattern of ten letters is observed. Here, ten successive time bins for a given neuron give a word. The results are shown in Fig. 1. Compare the performance of the two models, and conclude about the importance of covariances in this system.

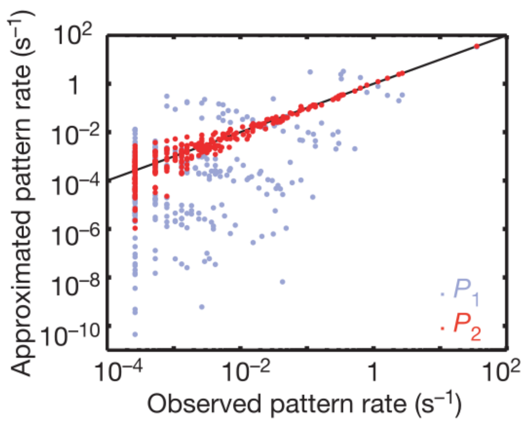

Figure 1: **Quality of predictions from maximum entropy inference.** The rate of occurrence of each firing pattern predicted from the maximum entropy model that takes into account means and covariances (red) and just means (blue) are plotted against the measured rate. The black line shows equality. *Reproduced from Ref. [1].*

References

[1] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, Apr 2006.