Solutions 11: Maximum entropy

BIO-369

Prof. Anne-Florence Bitbol
EPFL

1 Some examples

a) We have measured the average value of X, obtaining m. We need to find a distribution that satisfies the constraints

$$\sum_{x=0}^{\infty} P(x) = 1\,, \tag{1}$$

and

$$\sum_{x=0}^{\infty} xP(x) = m\,. \tag{2}$$

We thus need to maximize the function

$$\tilde{H}(X) = -\sum_{x=0}^{\infty} P(x)\log_2 P(x) + \frac{\lambda}{\ln(2)}\left[1 - \sum_{x=0}^{\infty} P(x)\right] + \frac{\mu}{\ln(2)}\left[m - \sum_{x=0}^{\infty} xP(x)\right]\,, \tag{3}$$

which gives for all non-negative integer value x:

$$P(x) = \frac{e^{-\mu x}}{e^{1+\lambda}}\,. \tag{4}$$

Next, we need to use the constraints in Eqs. 1 and 2 to find the values of λ and μ. Using Eq. 4, and recognizing a geometric series (and assuming $\mu > 0$), Eq. 1 becomes

$$\sum_{x=0}^{\infty} e^{-\mu x} = \frac{1}{1 - e^{-\mu}} = e^{1+\lambda}\,, \tag{5}$$

and we can re-express Eq. 2 using

$$-\sum_{x=0}^{\infty} xe^{-\mu x} = \frac{\partial}{\partial \mu}\sum_{x=0}^{\infty} e^{-\mu x} = \frac{\partial}{\partial \mu}\left(\frac{1}{1 - e^{-\mu}}\right) = \frac{-e^{-\mu}}{(1 - e^{-\mu})^2}\,, \tag{6}$$

obtaining

$$\frac{e^{-\mu}}{(1 - e^{-\mu})^2} = m\, e^{1+\lambda}\,. \tag{7}$$

Combining Eqs. 5 and 7 gives

$$\frac{e^{-\mu}}{1 - e^{-\mu}} = m\,, \tag{8}$$

and thus

$$\mu = \log\left(\frac{m+1}{m}\right)\,. \tag{9}$$

Next, Eq. 5 gives

$$\lambda = \log(m+1) - 1\,. \tag{10}$$

Injecting these expressions in Eq. 4 finally yields, for all non-negative integer value x:

$$P(x) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^x\,. \tag{11}$$

b) For a continuous random variable X, the maximum-entropy probability density $p(x)$ takes the same form as $P(x)$ in the discrete case. Thus, with the constraint of the measured mean m, we have as in Eq. 4, for all $x \geq 0$:

$$p(x) = \frac{e^{-\mu x}}{e^{1+\lambda}} \,. \tag{12}$$

In the continuous case, the normalization and mean constraints read

$$\int_0^\infty p(x)dx = 1 \,, \tag{13}$$

and

$$\int_0^\infty xp(x)dx = m \,. \tag{14}$$

Combining Eqs. 12 and 13 and performing the integral yields

$$\mu e^{1+\lambda} = 1 \,, \tag{15}$$

and similarly, combining Eqs. 12 and 14 and performing the integral yields

$$\mu^2 e^{1+\lambda} = \frac{1}{m} \,. \tag{16}$$

Then, combining Eqs. 15 and 16 gives

$$\mu = \frac{1}{m} \,, \tag{17}$$

and

$$\lambda = \log(m) - 1 \,. \tag{18}$$

Injecting these expressions in Eq. 12 finally yields, for all $x \geq 0$:

$$p(x) = \frac{1}{m} \exp\left(-\frac{x}{m}\right) \,. \tag{19}$$

This is the exponential distribution with mean m.

c) We now have the following three constraints:

$$\int_{-\infty}^\infty p(x)dx = 1 \,, \tag{20}$$

and

$$\int_{-\infty}^\infty xp(x)dx = m \,, \tag{21}$$

and

$$\int_{-\infty}^\infty x^2 p(x)dx = q \,. \tag{22}$$

According to the usual maximum entropy construction extended to continuous variables, we expect the form

$$p(x) = \frac{e^{-\mu x - \nu x^2}}{e^{1+\lambda}} \,. \tag{23}$$

This corresponds to a Gaussian distribution. Due to the constraints, it needs to have a mean m and a variance $V = q - m^2$, and to be normalized. Thus, it can be recast as

$$p(x) = \frac{\exp\left[-\frac{1}{2V}(x-m)^2\right]}{\sqrt{2\pi V}} \,. \tag{24}$$

Note that we can also find this form directly by imposing the three constraints and calculating the values of μ, ν and λ.

2 Maximum entropy modeling of neuroscience data

a) See Jupyter notebook.

b) See Jupyter notebook. In the scatter plot, we can see that the level of activity of this neuron is quite heterogeneous in time, with times during the movie when it is quite inactive and others when it is quite active. Conversely, we can also see that the neuron activity is rather stereotypical at a given time of the movie, but not perfectly. Thus we observe vertical bands of activity but with some fluctuations. The heterogeneity in the timing of spikes across replicates of the movie is more apparent when zooming to see individual markers corresponding to spikes.

c) See Jupyter notebook. The second neuron in the dataset has a lower level of activity overall than the first one, but otherwise the observations are similar.

d) See Jupyter notebook.

e) See Jupyter notebook. It happens that two spikes may occur in the same time bin, as the maximum "frequency" is larger than one. To correct this, we replace by one the frequencies that are larger than one. Note that we could deal with this more precisely but it remains rare, so this correction suffices.

f) We now make the mapping that a value 1 is 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 $\{-1, 1\}$. Denoting by P_i the probability that neuron i is active, then the probability that it is inactive is $1 - P_i$ and thus the mean of X_i reads

$$\langle x_i \rangle = P_i \times 1 + (1 - P_i) \times (-1) = 2P_i - 1 \,. \tag{25}$$

See Jupyter notebook.

g) See Jupyter notebook. The mean activities are quite small, close to the minimum possible value of -1. This reflects the fact that neurons are inactive most of the time. The scatter plot of these mean activities shows that they are quite heterogeneous. We also notice that the first and second neurons analyzed at the beginning of the problem were indeed quite strongly different, with the first one being among the most active ones and the second one among the least active ones.

h) See Jupyter notebook. The largest value of covariance is in the diagonal (it's a variance), and thus setting the diagonal to zero allows to better visualize the covariances between different neurons. They are quite heterogeneous, and most of them are positive.

Correlation coefficients are nice because they are normalized. Because of this normalization, the diagonal is full of ones, and it is also better to set it to zero to better visualize the rest. We observe that the normalization substantially changes the picture. But correlations are also quite heterogeneous. The mean correlation coefficient between two different neurons is 0.082, which is quite small, and seems to indicate that correlations are not very important in this system.

i) The form of the maximum entropy distribution consistent with the measurements of the mean activity of each neuron and of their covariances resembles those we are used to, with an exponential form and parameters in front of the functions of the random variables that are being constrained by the measurements. However, it is more general, since we are now modeling the joint probability of N random variables $(x_1, x_2, \ldots, x_N)$ and not that of one random variable. Here we see that the formula involves only terms that depend on one or two neurons in the exponential. This is because we have constrained terms that depend only on one or two neurons, the mean activity of each neuron and the covariances. As a side remark, note that in statistical physics, this formula corresponds to the Boltzmann distribution associated to an Ising model with fields h_i and couplings J_{ij}.

See Jupyter notebook. The parameters J_{ij} are also quite heterogeneous, and they are quite different from the covariances, and they are mainly negative. They also have quite large absolute values (absolute values much larger than one appear), which means that the approximation of small couplings used here is not well justified and that we would need to go beyond this to have a good estimate of the maximum entropy distribution for our neuron system.

j) By analogy with the rest, the maximum entropy distribution consistent with the measurements of the mean activity of each neuron would read

$$P(x_1, x_2, \ldots, x_N) = \frac{1}{Z} \exp\left[-\sum_{i=1}^{N} h_i x_i\right] , \tag{26}$$

where Z is a normalization constant.

The figure shows that the model that constrains the means and the covariances of neuron activities makes predictions that are much closer to the actual measurements than the model that only constrains the mean activities. This means that the covariances actually matter for this system, despite the small average value of the correlation coefficients seen above. The behavior of these neurons cannot be well approximated as that of a set of independent neurons, interactions between them matter.