Given sample, infer population

Luca <piero.mana@ntnu.no>

Draft of 29 January 2019 (first drafted 25 January 2019)

An analysis of the problem of inferring the state of a population of neurons from that of a sample.
Note: Dear Reader & Peer, this manuscript is being peer-reviewed by you. Thank you.

1 The question

The problem we want to consider is the inference about the state of a population of neurons from the observation of the state of a sample of that population. By inference we mean *the numerical evaluation of our degree of belief*, in this case about the population's state. The state is taken to be the binarized activity from a time-binned sequence. Denote by N the size of the population, by n that of the sample, by $S_i(t)$ the activity of the ith neuron in the population at time t, by $\boldsymbol{S}(t) \coloneqq \bigl(S_1(t), \dots, S_N(t)\bigr)$ the joint activity – the state – of the population, and by $s_j(t)$, $\boldsymbol{s}(t) \coloneqq \bigl(s_1(t), \dots, s_n(t)\bigr)$ the corresponding activities and state of the neurons making up the sample. We will discuss the exact relationship between $\boldsymbol{S}$ and $\boldsymbol{s}$ in the next sections. Our main task is to calculate our degree of belief

$$\mathrm{p}[\boldsymbol{S}(t) \mid \boldsymbol{s}(t), I], \tag{1}$$

where I denotes other initial information and assumptions. Many points of our discussion apply to more general definitions of 'state'.

In order to better understand what our inference is about, let's also stress what it is *not* about. Our inference is not about the *dynamics* of the population. This latter inference is roughly as follows. We assume that the state $\boldsymbol{S}(t)$ at time t is determined through a dynamical law by the states $\{\boldsymbol{S}(\tau)\}_{\tau<t}$ at some previous times together with some external quantities $Q(t)$ (such as physical states of synapses, inputs from peripheral nervous system, and similar extra-neuronal quantities):

$$\boldsymbol{S}(t) = F[\{\boldsymbol{S}(\tau)\}_{\tau<t}, Q(t)]. \tag{2}$$

We are uncertain about the mathematical form of the dynamical law F and the values of the external quantities $Q(t)$. We can therefore consider various degrees of belief: for example the one about $S(t)$ given only knowledge about some previous states:

$$\mathrm{p}[S(t) \mid \{S(\tau)\}_{\tau<t}, I], \tag{3}$$

or the one about the dynamical law, given a time sequence of states:

$$\mathrm{p}[F \mid \{S(\tau)\}, I]. \tag{4}$$

Our present problem doesn't concern this kind of inferences, but it's very relevant to them: to infer the dynamics, eq. (4), we usually must first infer the states $S(\tau)$ from the observation of a population sample.

Note that if the dynamics is excluded from our problem, then samples at times $\tau < t$ cannot be used for the inference of the population state at time t, because such inferential chain involves the dynamics: schematically, the inference would be $s(\tau) \rightsquigarrow S(\tau) \rightsquigarrow S(t)$, and the latter step involves the degree of belief (3). Our discussion will therefore refer to one time t only, conveniently suppressed from our notation.

To calculate our degree of belief (1), the probability calculus requires us to specify: 1. our initial belief distribution about the population state:

$$\mathrm{p}(S \mid I); \tag{5}$$

2. our belief distribution about the sample state given the population state:

$$\mathrm{p}(s \mid S, I), \tag{6}$$

which we can call the 'sampling distribution'. The two distribution above yield the distribution (1) by Bayes's theorem:

$$\mathrm{p}(S \mid s, I) = \frac{\mathrm{p}(s \mid S, I)\, \mathrm{p}(S \mid I)}{\sum_S \mathrm{p}(s \mid S, I)\, \mathrm{p}(S \mid I)}. \tag{7}$$

Let's investigate the sampling distribution (6). First of all we note that we may *label* the N neurons in an arbitrary way – this doesn't mean that we consider them identical or indistinguishable. It is then convenient to give the labels $1, \dots, n$ to the neurons we have measured, and the remaining $N - n$ labels to the rest. Then we have the identity

$$s_i = S_i, \quad i \in \{1, \dots, n\}, \tag{8}$$

and the sampling distribution (6) is a delta:

$$\mathrm{p}(\boldsymbol{s} \mid \boldsymbol{S}, I) = \prod_{i=1}^{n} \delta(s_i, S_i). \tag{9}$$

Let's now investigate our initial belief (5) about the population state. To start with, I'd like to consider belief distributions of a maximum-entropy form and offer a couple of comments on them, because they seem very popular in the literature.

If the activities of the neurons are binarized, the set of all possible population states $\{\boldsymbol{S}\}$ is discrete, of cardinality 2^N. The set of all possible initial belief distributions for $\boldsymbol{S}$ has then dimension $2^N - 1$ because of normalization. It is a simplex. Each such distribution has moments – for example, $\mathrm{E}(S_3 \;\; S_5 \;\; S_8 \mid \; I)$ — with precise numerical values. A maximum-entropy distribution is chosen by first choosing a subset of distributions having specific values for some moments, and then selecting the distribution having maximum Shannon entropy in this subset. Such a distribution is unique because the fixed-moment subsets are convex and the Shannon entropy is a convex function. A maximum-entropy distribution is therefore identified by the moments chosen – for example, first and second moments – and their numerical values. We can write this as

$$\mathrm{p}(\boldsymbol{S} \mid \boldsymbol{m}, I_{\mathrm{ME}}), \tag{10}$$

a familiar example being

$$\mathrm{p}(\boldsymbol{S} \mid \{m_i, m_{ij}\}, I_{\mathrm{ME}}) = \\ \frac{1}{Z(\{m_i, m_{ij}\})} \exp\Bigl[\sum_i h_i(\{m_i, m_{ij}\})\, S_i + \sum_{i,j}^{i<j} J_{ij}(\{m_i, m_{ij}\})\, S_i\, S_j\Bigr], \tag{11}$$

where $\{m_i\}$ are the N first moments, $\{m_{ij}\}$ the $\binom{N}{2}$ second moments, Z is a normalization constant, and $\{h_i, J_{ij}\}$ are specific one-one functions of the moments.

This kind of distributions can assign asymmetric degree of belief about the activities of the neurons – for example a higher belief that neuron 14 is active than that neuron 6 is active. By keeping the *kind* of moments (for example, all first moments and some specific third moments) fixed but choosing different numerical values for them we form a set of distributions. If the number of moments considered is less than $2^N - 1$ then this set has strictly lower dimension than the simplex.

A remark may be useful speaking of maximum-entropy distributions. With other kinds of inference, convex mixtures of maximum-entropy distributions are sometimes considered, which can be written as

$$\mathrm{p}(\boldsymbol{S} \mid I_{\mathrm{ME}}) = \int \mathrm{d}\boldsymbol{m}\, \mathrm{p}(\boldsymbol{S} \mid \boldsymbol{m}, I_{\mathrm{ME}})\, \mathrm{p}(\boldsymbol{m} \mid I_{\mathrm{ME}}). \tag{12}$$

For our present inference, however, such a mixture is not meaningful because redundant: its redundancy is clear when we consider the distribution for $\boldsymbol{m}$ conditional on perfect knowledge of the state $\boldsymbol{S}$:

$$\mathrm{p}(\boldsymbol{m} \mid S, I_{\mathrm{ME}}) \propto \mathrm{p}(\boldsymbol{S} \mid \boldsymbol{m}, I_{\mathrm{ME}})\, \mathrm{p}(\boldsymbol{m} \mid I_{\mathrm{ME}}). \tag{13}$$

This distribution is not a delta, that is, it says we're uncertain about the values of the parameter $\boldsymbol{m}$ – even though the state $\boldsymbol{S}$ is known!

The reason for this redundancy is that two different sets of weights $\{\mathrm{p}(\boldsymbol{m} \mid I'_{\mathrm{ME}})\}$, $\{\mathrm{p}(\boldsymbol{m} \mid I''_{\mathrm{ME}})\}$ for $\boldsymbol{m}$ may yield the same resulting distribution for $\boldsymbol{S}$. Mixtures of maximum-entropy distributions like (12) make sense when we are making inferences about an *unlimited* sequence of states – for example, a time sequence $\big(\boldsymbol{S}(t_1), \boldsymbol{S}(t_2), \dots\big)$ – because in this case each set of weights for $\boldsymbol{m}$ gives rise to a unique distribution *for the unlimited sequence*.

It's important to ask ourselves: in which experimental situations do asymmetric belief distributions represent our initial state of knowledge about the population state? The signals of the neurons picked up by neural recording instruments is usually