Given sample, infer population

Luca <piero.mana@ntnu.no>

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 $\boldsymbol{S}(t) \coloneqq \big(S_1(t), \dots, S_N(t)\big)$ the joint activity – the state – of the population, and by $s_j(t)$, $\boldsymbol{s}(t) \coloneqq \big(s_1(t), \dots, s_n(t)\big)$ 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

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

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:

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

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

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

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:

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

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

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

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

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

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 $1, \ldots, n$ to the neurons we have measured, and the remaining $N - n$ labels to the rest. Then we have the identity

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

and the sampling distribution (6) is a delta:

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

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 $\{S\}$ is discrete, of cardinality 2^N. The set of all possible initial belief distributions for S has then dimension $2^N - 1$ because of normalization. It is a simplex. Each such distribution has moments – for example, $E(S_3 \ S_5 \ S_8 \mid I)$ — 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

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

a familiar example being

$$p(S \mid \{m_i, m_{ij}\}, I_{\mathrm{ME}}) =$$

$$\frac{1}{Z(\{m_i, m_{ij}\})} \exp\left[\sum_i h_i(\{m_i, m_{ij}\}) S_i + \sum_{i,j}^{i<j} J_{ij}(\{m_i, m_{ij}\}) S_i S_j\right], \tag{11}$$

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

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

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 S:

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

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 S is known!

The reason for this redundancy is that two different sets of weights $\{p(\boldsymbol{m} \mid I'_{\text{ME}})\}, \{p(\boldsymbol{m} \mid I''_{\text{ME}})\}$ for $\boldsymbol{m}$ may yield the same resulting distribution for 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 $(S(t_1), S(t_2), \dots)$ – because in this case each set of weights for $\boldsymbol{m}$ gives rise to a unique distribution *for the unlimited sequence*. In fact, in this case such mixtures are called models by sufficient statistics (Bernardo et al. 2000 § 4.5).

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 are usually

2   Intro

The probabilistic study of the activity of networks of neurons has enjoyed many mathematical advances from other scientific fields, statistical

mechanics in particular. Together with their mathematical understanding, these fields have also brought into neuroscience their technical terminology and, behind it, specific ways of physically picturing the problem.

Although the mathematical tools are a great help, I believe that the technical jargon and physical picture of other fields will in the long run be a hinder to understanding neuronal-network activity, because the physical phenomena at its core are very different from those of these fields.

Some people will object that this is 'just semantics' and it's unimportant. People who make this kind of objection are usually poorly acquainted with the history of science, where most of the advances have been, first of all, advances in semantics. The mathematical Ptolemaic system, for example, had great predictive power and could be improved without limit. The reason for its ultimate dismissal in favour of the Copernican-Keplerian-Newtonian one was that the latter was based on a more fruitful physical *picture*, which in turn led to more agile mathematics (Kline 1982 ch. ?). Another example is semantic change about the idea of pressure – pressure as something we can speak about even when there are no physical walls – offered by Euler (Euler 1757; Truesdell 1954a; 1968 ch. IV), which eventually led to continuum thermomechanics. The most famous example is Einstein's (1905) semantic analysis of the concept of 'simultaneity', which led to a completely different understanding of the mathematically already known Lorentz transformations and their related phenomena of length contraction and time dilation (and of the relation between mass and energy, already pointed out by Poincaré 1900 p. 487). A final example is Born (1926a p. 865, footnote) ***

Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Bernardo, J.-M., Smith, A. F. (2000): *Bayesian Theory*, reprint. (Wiley, New York). First publ. 1994.

Born, M. (1926a): *Zur Quantenmechanik der Stoßvorgänge*. Z. Phys. **37**[12], 863–867. http://www.physics.princeton.edu/~mcdonald/examples/EM/. See also Born (1926b,c).

—— (1926b): *Quantenmechanik der Stoßvorgänge*. Z. Phys. **38**[11–12], 803–827. http://www.physics.princeton.edu/~mcdonald/examples/EM/. See also Born (1926a,c).

—— (1926c): *Zur Wellenmechanik der Stoßvorgänge*. Ges. d. Wiss. Nachrichten (Gött.) Math.-Phys. Kl. **1926**, 146–160.

Einstein, A. (1905): *Zur Elektrodynamik bewegter Körper*. Ann. der Phys. **17**, 891–921. Transl. in Einstein (1989a), Doc. 23, pp. 140–171.

—— (1989a): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909. (English translation)*. (Princeton University Press, Princeton). Transl. of Einstein (1989b) by Anna Beck and Peter Havas.

—— (1989b): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909*. (Princeton University Press, Princeton). Ed. by John Stachel. Transl. in Einstein (1989a).

Euler, L. (1757): *Principes généraux de l'état d'équilibre des fluides*. Mémoires de l'académie des sciences de Berlin **11**, 217–273, 402(plates). http://eulerarchive.maa.org/; presented 1753; repr. in *Opera Omnia* series 2, vol. 12, pp. 2–53.

Kline, M. (1982): *Mathematics: The Loss of Certainty*. (Oxford University Press, Oxford). First publ. 1980.

Poincaré, H. (1900): *La théorie de Lorentz et le principe de réaction*. Archives néerlandaises des Sciences exactes et naturelles **5**, 252–278. Transl. into English in http://physicsinsights.org/poincare-1900.pdf.

Truesdell III, C. A. (1954a): *Rational Fluid Mechanics, 1687–1765*. In: Truesdell (1954b), VII–CXXV. Editor's introduction to Euleri Opera omnia II 12.

—— ed. (1954b): *Leonhardi Euleri commentationes mechanicae ad theoriam corporum fluidorum pertinentes. Vol. I*. (Orell Füssli, Zurich).

—— (1968): *Essays in the History of Mechanics*. (Springer, Berlin).