Open Science Framework DOI:10.31219/osf.io/***

Memos on inferring connectivity [draft]

P.G.L. Porta Mana
<piero.mana@ntnu.no>

B. Jacobsen
<bente.jacobsen@ntnu.no>

17 October 2019; updated 22 January 2020

Some memos on the problem of inferring connectivity by means of the probability-calculus aka Bayesian theory.

1 Introduction

We are interested in quantifying the proportion with which several brain regions project to others. It's impossible, at least at present, to simply count individual projections. What's done experimentally, with very clever techniques, is analogous to survey sampling: we manage to see, more or less clearly, a small sample of projections. From this sample we want to infer the larger picture, for a single brain and on average for a 'typical' brain. We show here how to make such inferences using the probability-calculus[1], also known as Bayesian theory.

We accompany the method with an extensive discussion of its steps and assumptions, for the sake of researchers unfamiliar with it. But we believe that some of the points of view will be of interest also to experts in the method.

2 Synopsis of the problem

We're considering two sets of neurons, which we call *target* set and *input* set. The neurons in the target set are of several classes, for example parvalbumin or somatostatin interneurons; and are distributed across several regions, for example medial septal complex or thalamus). Same holds for the input set. In general we are interested in the number of projections of input neurons to target neurons, and in how such projections are distributed across regions, or classes, or both, of neurons. More specifically we focus here on the following kind of question: How many input neurons in a region R (independently of their class) project to target neurons of class C (in a specific location)? And we want to

[1] Jaynes 2003; Hailperin 1996; Sox et al. 2013.

compare the answer for several combinations of input regions R and target classes C, thus obtaining the proportion for each combination.

But the question above is still imprecise. Are we asking about a specific mouse? or about the average across all mice having some specific characteristics? These two cases can have slightly different answers, and their difference is important for the application of the probability-calculus and for experimental verification. We'll focus on these two kinds of question:

Individual: If we were able to perfectly count, in the next mouse to be examined, the number of input neurons in region R projecting on average to a target neuron of class C, – what number would we find?

Collective or average: If we were able to perfectly count, for each mouse (with some specific traits such as age or gender and so on) to be examined in the future or that could have been examined in the past, the number of input neurons in region R projecting to target neurons of class C, – what average number would we find?

It's intuitively clear that the answer to the second question comes from the combination of answers to the first, for many mice. In either case, our questions are about *statistics*: total numbers, averages, deviations from averages, and so on.

The observations and data we use: In a mouse, virus methodologies allow us to select several neurons in the target set and to collectively observe which neurons in the input set project to them. We can't distinguish, however, which input neuron projects to which target neuron. Such observations are made in several mice.

There's a complication: to make this kind of observation the mouse must be killed. If it's killed too early, the virus hasn't had the time to spread to all input neurons. If the mouse is killed too late, the starter neurons decay and are no longer visible. So we also need to guess, for each observed mouse, whether there were additional starter neurons, no longer visible, and what proportion of input neurons have been reached by the virus. As data we have the number of days between virus injection and killing, for each mouse. This complication for the moment is set aside to simplify the method, but will be taken into account later.

Now let's sketch a schematic picture of the strategy to answer the two questions above.

3 Formulation and strategies

3.0 Preliminary remarks

Probabilistic inferences are just generalizations of logical inferences. Their analysis therefore involve a logical analysis of the problem. A probabilistic inference generally boils down to these steps:

1. Find a set of simple, factual statements, in terms of which the questions and the data can be formulated. These are called 'atomic statements'; atomic in the sense that we won't analyse them into more precise statements. The choice of atomic statements is not unique and is mostly dictated by convenience.
2. Express the data and the possible answers to the questions in terms of the atomic statements:
$$\begin{aligned} \text{data} &= \textit{[some combination of atomic statements]}, \\ \text{answers} &= \textit{[some combination of atomic statements]}. \end{aligned} \tag{1}$$
3. Assign a joint probability distribution for all the possible combinations of these atomic statements,
$$\mathrm{p}(\textit{[combination of atomic statements]} \mid B). \tag{2}$$
This probability distribution is called the 'pre-data' or 'prior' distribution, because it depends on our background information and assumptions, denoted by B, but not yet on the experimental data.
4. Use the probability-calculus to determine, from the pre-data distribution (2) and the expressions (1), the probability for each answer of interest, conditional on the data observed and on the pre-data information:
$$\mathrm{p}(\text{answer} \mid \text{data}, B) = \frac{\mathrm{p}(\text{answer}, \text{data} \mid B)}{\mathrm{p}(\text{data} \mid B)}. \tag{3}$$
The probabilities for all possible answers form the so-called 'post-data' or 'posterior' distribution.

Why do we need a pre-data distribution? can't we find the final probability from the data alone, without background information and assumptions? – No, we can't; it's a logical impossibility. Just as in the truth-calculus (formal logic) we can't prove any theorem if we don't specify some initial axioms or postulates[2], so in the probability-calculus we can't

[2] Barwise et al. 2003 § 2.1 p. 43.

derive any probabilities without some initial 'probabilistic postulates'[3]. *All* statistical and probabilistic methods use assumptions.

These steps are very interdependent and usually not faced sequentially. A choice of atomic statements, for example, may lead to difficulties in the assignment of the pre-data probabilities. Trying to simplify the latter probabilities may lead us back to consider an alternative choice of atomic statements.

The first and third steps always involve some degree of simplification and abstraction of our problem, necessary for eliminating features of the problem that are unimportant, and for making the problem mathematically and computationally tractable. The third step often involves some mathematical approximations.

We now discuss these steps for our problem.

3.1 Step 1: atomic statements

In our case there are several possibilities for this first step, the choice of atomic statements. Our choice below tries to simplify the problem, to make it computationally tractable, without making it unrealistic.

We consider a very large number of mice: those we have observed and all those that we might observe in the future or we could have observed in the past. For our inferences to be valid, these mice must be judged 'similar' in some respects. We label each mouse with an index $m = 1, 2, \dots$.

For the moment *let's fix a specific input region*, for example the medial septal complex or thalamus, *and a specific target class*, for example parvalbumin interneurons, in a specific target region. We'll discuss later how to combine inferences about several input regions and target classes.

For each mouse, we consider all neurons constituting the chosen input region, labelling them $i = 1, 2, \dots, I^{(m)}$, and all neurons constituting the target class, labelling them $t = 1, 2, \dots, T^{(m)}$. The total numbers of these neurons can be different from mouse to mouse, hence the '(m)' subscript, which we shall use also for other quantities that may vary from mouse to mouse.

The values of the three labels i, t, m may be arbitrary or carry some information. For example, m may reflect a temporal order, and i, t may

[3] Hailperin 2011; Johnson 1924 p. 182.

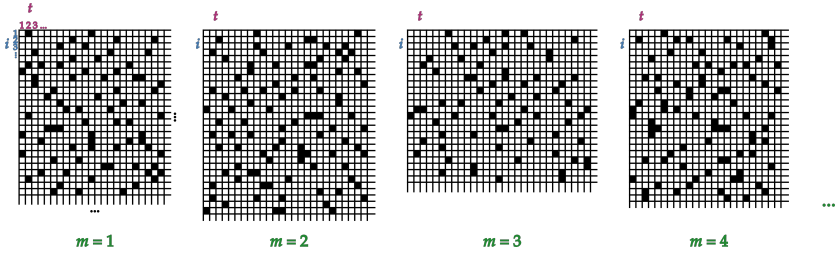

Figure 1 Imaginary layout of the connectivity from input to target region for all mice

each reflect a topographical order, for example ventral to dorsal. In the next step, however, we'll assume such order to be unimportant.

Now that we've labelled the mice and the input and target neurons of interest, let's consider some simple, specific statements about projections. for example:

> *In mouse $m = 3$, the neuron $i = 5$ in the input region projects to the neuron $t = 4$ of the target class.*

Such a statement must be either true or false, even if we may not have the experimental technology to look and see. If we could experimentally ascertain each of these statements, for all mice m, input neurons i, target neurons t, then our problem would be solved. We'd just have to count, and we could report the exact projection proportions for each mouse, and also averages across all mice. *We'll use this kind of statements as our atomic statements.*

The situation collectively described by these statements is visualized in fig. 1. We plot a grid for each mouse m. Each row i in the grid represents a neuron in the input region, each column t a neuron of the target class. If in mouse m neuron i projects to neuron t, then we colour the corresponding entry (i, t) in grid m in black; or we can write down the number 1 in it. If it doesn't project, then we use white or the number 0.

Let's indicate the entries of all grids with the binary indicators $c_{it}^{(m)}$. If input neuron i projects to target neuron t in mouse m, we write $c_{it}^{(m)} = 1$; otherwise we write $c_{it}^{(m)} = 0$. Then *our atomic statements will be of the form* '$c_{it}^{(m)} = 0$' *or* '$c_{it}^{(m)} = 1$' for all m, i, t we need.

By assigning a probability distribution for all combinations of these indicators,

$$\mathrm{p}(\{c_{it}^{(m)}\} \mid B) \equiv \mathrm{p}\big(c_{11}^{(1)}, c_{12}^{(1)}, \dots, c_{21}^{(1)}, c_{22}^{(1)}, \dots,\ c_{11}^{(2)}, \dots \mid B\big), \tag{4}$$

we can calculate from it the probabilities (3) for the answers to our main questions.

First let's see how to express those answers and our data in terms of the atomic statements.

3.2 Step 2: data and answers in terms of atomic statements

3.2.1 *Our questions*

Let's start with the answers to our questions. As discussed in § 2, our questions are about individual and collective statistics. Let's then introduce some statistics, expressed in terms of the atomic statements.

Consider first a specific mouse m and target neuron t. A useful statistics is the total number of input neurons that project to that target neuron. Denote this by $N_t^{(m)}$. It is of course given by

$$N_t^{(m)} \coloneqq \sum_i c_{it}^{(m)}, \tag{5}$$

that is, by counting the black squares in the tth column of the mth grid in fig. 1.

This statistic is useful because from it we can calculate, for example, the average number of projections to the target neurons in that mouse:

$$\overline{N}^{(m)} \coloneqq \frac{1}{T^{(m)}} \sum_t N_t^{(m)}, \tag{6}$$

and the variance from that average,

$$\frac{1}{T^{(m)}} \sum_t \big(N_t^{(m)} - \overline{N}^{(m)}\big)^2, \tag{7}$$

which answer some basic questions about a specific mouse.

But we may be interested in a more statistics of the projections in that mouse. For example, how many target neurons receive no projections from the input region? how many receive only one projection? how many receive two? And so on. This statistics is given by a histogram

$\boldsymbol{H}^{(m)} \coloneqq \left(H_0^{(m)}, H_1^{(m)}, H_2^{(m)}, \dots, H_{I^{(m)}}^{(m)}\right)$ of the number of projections for that mouse: $H_0^{(m)}$ is the number of neurons receiving no projections, $H_1^{(m)}$ the number receiving only one projection, and so on, up to the total number of neurons $I^{(m)}$ in the input region. Mathematically this histogram is given by

$$H_n^{(m)} \coloneqq \textstyle\sum_t \delta\left(N_t^{(m)} = n\right) \tag{8}$$

where δ is the Kronecker delta. From such histogram we can calculate the mean and variance above; for example $\overline{N}^{(m)} = \frac{1}{I^{(m)}} \sum_n n\, H_n^{(m)}$.

Also useful is the normalized version of the histogram, $\boldsymbol{F}^{(m)} \coloneqq \boldsymbol{H}^{(m)}/T^{(m)}$, corresponding to the relative frequency distribution of projections.

At the level of the whole mouse population (or subpopulation with specific traits, as discussed before), we are interested again on the average number of projections from the input region to an target neuron, and so on. But we can ask for more statistical information. For example, if we had the histograms $\boldsymbol{F}^{(m)}$ for each mouse m, we could then ask about the average histogram of projections $\overline{\boldsymbol{F}} \coloneqq \frac{1}{M} \sum_m \boldsymbol{F}^{(m)}$ across the whole population, where M is the total number of mice. We could also ask about the variance from such an average histogram. More in detail, we could build a 'super-statistics' or 'histogram of histograms': we imagine to check every mouse, and for each we write down its histogram of projections $\boldsymbol{F}^{(m)}$; then we count how many mice have the particular histogram $\boldsymbol{F}'$ in common, how many have another particular histogram $\boldsymbol{F}''$ in common, and so on. We obtain a 'super-histogram' $\boldsymbol{S} = (S_{\boldsymbol{F}})$: $S_{\boldsymbol{F}'}$ is the proportion of mice that have the same histogram $\boldsymbol{F}'$, and so on. Mathematically it's expressed as

$$S_f \coloneqq \frac{1}{M} \textstyle\sum_m \delta\left(\boldsymbol{F}^{(m)} = f\right), \tag{9}$$

similarly to the expression for the individual statistics (8). It allows us to say what's the most common histogram of projections in our mice population, what's the average histogram, and how much the histograms can vary from the average one, across the population. Since each possible $\boldsymbol{F}^{(m)}$ is a tuple with $I^{(m)} + 1$ elements, the distribution $\boldsymbol{S}$ is over an $(I+1)$-dimensional space, where I is the largest of all $I^{(m)}$.

3.2.2 *Our data*

The expression of the data in terms of the atomic statements $\{c_{it}^{(m)}\}$ requires some more thinking. We're still focusing on a specific input region (say, thalamus) and a specific target class (say, parvalbumin). Our data consists, for a number of mice, of the total number of target neurons marked by the virus, and the total number of input neurons that have been infected by the target ones, and thus also marked by the virus. Let's use a circumflex to indicate observed quantities. The observed mice are indexed with $\hat{m}$, the marked target neurons with $\hat{t}$, and the marked input neurons with $\hat{\imath}$. The total numbers of marked target and input neurons in mouse $\hat{m}$ are $\hat{T}^{(m)}$ and $\hat{I}^{(m)}$.

Consider a specific mouse (we omit its index $\hat{m}$ for the moment), and the grid that represents its projections. Owing to our assumption that the specific order and indexing of target and input neurons doesn't matter, we can rearrange the grid so that the input neurons marked by the virus are the first $\hat{I}$ rows, and the marked target neurons are the first $\hat{T}$ columns; see fig. 2.

Because of the way the virus infection works, our data correspond to the following two main pieces of information; refer to the grid of fig. 2:

- α: Each marked input neuron must project to *at least one* marked target neuron – otherwise it wouldn't have been infected and marked. This means that *each row* in the upper-right, red-bounded part of the grid *must have at least one black entry* – but we don't know how many black entries and where.
- β: The unmarked input neurons *cannot* project to the marked target neurons – because otherwise they would have been infected and would be marked instead. Thus the lower-right, blue-bounded part of the grid has no black entries.

On the other hand we don't know anything about the projections of marked and unmarked input neurons to the unmarked target neurons. This is represented by the yellow-bounded region on the right with the large yellow question mark.

The expression of β is simple: it corresponds to

$$\beta: \qquad c_{it} = 0 \quad \text{for } i = \hat{I}+1, \dots, I \quad \text{and } t = 1, \dots, \hat{T}. \tag{10}$$

The expression of α is slightly more involved. Fix a row i; the statement that this row is empty is '$c_{it} = 0$ for $t = 1, \dots, \hat{T}$'. Such a

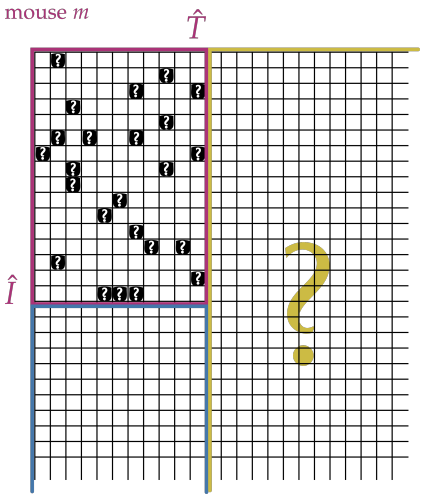

Figure 2 Imaginary layout of the connectivity data for an observed mouse

statement must be false (because a row must contain at least one black entry), whether we choose the row $i = 1$, or $i = 2$, and so on to $i = \hat{I}$. Then

$$\begin{aligned} \alpha: \quad & \text{not}\,\big[(c_{1t} = 0 \quad \text{for } t = 1,\dots,\hat{T}) \\ & \text{or}\,(c_{2t} = 0 \quad \text{for } t = 1,\dots,\hat{T}) \\ & \text{or}\,\cdots \\ & \text{or}\,\big(c_{\hat{I}t} = 0 \quad \text{for } t = 1,\dots,\hat{T}\big)\big], \end{aligned} \tag{11}$$

which can be more compactly written using Boolean logical operators as

$$\alpha: \quad \neg \bigvee_{i=1}^{\hat{I}} \big[\bigwedge_{t=1}^{\hat{T}} (c_{it} = 0)\big]. \tag{12}$$

The statements α and β above refer to a specific mouse m. To express our data about all mice we simply 'and' these statements for all m.

Now that our data and the answers to our questions have been expressed as combinations of atomic statements, we turn to the pre-data probability distribution of the latter.

3.3 Step 3: pre-data probabilities for the atomic statements

We must now assign the pre-data probabilities

$$\mathrm{p}(\{c_{it}^{(m)}\} \mid B) \qquad (4)_{\mathrm{r}}$$

for all possible combinations of any number of atomic statements $c_{it}^{(m)}$.

This probability distribution represents our pre-data guess about the numbers presence or absence of projections. It should be a reasonable, unbiased guess.

Some researchers feel uncomfortable in quantifying such a guess. But, as we stressed in § 3, such a quantification is unavoidable: without it our experimental observations could not be probabilistically generalized to other neurons or mice. Some researchers also believe that there are statistical methods, such as so-called frequentist ones, which avoid this kind of pre-data assumptions. This is a mistaken belief. Those methods *do* make pre-data assumptions, but misleadingly keep them hidden. As I. J. Good said: 'The main difference is that in a non-Bayesian analysis more is swept under the carpet. This makes non-Bayesian methods politically expedient. The Bayesian is forced to put his cards on the table instead of up his sleeve. He thus helps others to improve his analysis, and this is always possible in a problem concerning the real world'[4] (Good was Turing's collaborator in the breaking of the *Enigma* code during World War II, using probabilistic methods). In fact, as the amount of experimental data increases, the effect of our pre-data guesses usually decreases. The Bayesian approach allows us to quantify this dependence and thus to decide whether the data we have collected are enough or whether we want to collect more.

To assign this distribution we are going to use three *symmetry* assumptions, at three increasing levels of generality: about the statistics $N_t^{(m)}$ for each individual target neuron of each individual mouse; about the statistics $\boldsymbol{F}^{(m)}$ for each individual mouse; and finally about the super-statistics $\boldsymbol{S}$ for all mice.

3.3.1 $c_{it}^{(m)}$ *for one target neuron and mouse: exchangeability*

Let's start by focusing on a specific but arbitrary target neuron in a specific but arbitrary mouse m. We'll suppress all '(m)' for the moment, to make the notation lighter.

Suppose we know the total number N_t of projections to that neuron from the input region, but not which neurons in the input neurons are projecting. We want to assign the probability distribution for the projection indicators c_{1t}, c_{2j}, and so on, conditional on knowledge of N_t. Our assumption is this: *there isn't any reason to believe that some input neuron i' projects to t, more than some other input neuron i''*. In other words,

[4] Good 1969 § 2.3 p. 26.

the specific index i of an shouldn't give us any special information about the projection. This is called an *exchangeability* assumption. In which cases can this assumption be at least approximately justified?

One case is if the indexing of the input neurons is done in a way completely unknown to us, for example by an automated unique-string generator. This isn't our case, because no matter how an input neuron is indexed we can always check what its location was. Another case is if all input neurons under consideration are very similar, and very close with respect to larger brain scales. Then none of them should find it easier to project than any other. A third case is if previous studies showed that the relative locations and morphological differences of the input neurons are unimportant.

The exchangeability assumption may therefore be unreasonable if the input neurons are located in widely different brain areas. In the latter case we can sort them into small neighbourhood (and morphology) groups, and apply the exchangeability assumption only within neighbourhoods. This is equivalent to redefining the input region R to a smaller one.

Exchangeability has a powerful mathematical consequence, called the de Finetti's representation theorem[5]. Effectively it's a combinatorial argument based on the assumed symmetry.

In our case, suppose we are interested in $n = n_0 + n_1$ input neurons, and we want the conditional probability that n_0 of them don't project to t, so that their indicators c_{it} are 0, and the other n_1 do project to t, and their indicators are 1. We know that there are I neurons in the whole input region, and that N_t of these project. By the assumed symmetry, this situation is equivalent to 'drawing without replacement', described by the hypergeometric distribution[6]:

$$\mathrm{p}(\{\underbrace{c_{\dots t} = 0}_{n_0 \text{ of them}}\}, \{\underbrace{c_{\dots t} = 1}_{n_1 \text{ of them}}\} \mid N_t, B) = \frac{1}{\binom{n}{n_0, n_1}} \binom{N_t}{n_1} \binom{I - N_t}{n_0} \Big/ \binom{I}{n}, \qquad (13)$$

where the parentheses on the right are binomial coefficients. This distribution simply comes from counting in how many ways we can draw n_0 white and n_1 black balls from an urn containing $I - N_t$ white and N_t black ones. The term $\binom{n}{n_0, n_1}$ is a multinomial coefficient, necessary

[5] De Finetti 1930; Hewitt et al. 1955; a more intuitive explanation is in Heath et al. 1976; for an informative review: Dawid 2013; the particular form used here is discussed in Diaconis 1977; Diaconis et al. 1980.

[6] Jaynes 2003 ch. 3.

because we are asking about the probability of a specific set of input neurons.

We have such distribution separately for each target neuron t. But the collection of these distributions depend on knowledge of all the $\{N_t\}$, which we don't know. We consider our guess about them next.

3.3.2 $N_t^{(m)}$ *for one mouse*

We're still focusing on a specific but arbitrary mouse m. Now we want to express our guesses about the numbers of projections N_t to a specific set of h target neurons.

Suppose we know the statistics $\boldsymbol{F} = (F_0, F_1, \dots, F_I)$, described in § 3.2.1, of such numbers of projections for all T target neurons. That is, F_0 is the proportion of all target neurons receiving no projections from the input region, F_1 the proportion receiving one projection, and so on. The number of target neurons in each case is given by $T\,F_n \equiv H_n$.

We make again an assumption of symmetry: *there isn't any reason for some target neuron j' to have more or less projections than some other target neuron j''*. This is again an assumption of exchangeability, and the discussion of its validity is analogous to that of the previous section.

The result for our guesses is also similar. Our situation in this case is analogous to 'drawing without replacement' from an urn having T balls of $I+1$ different colours or kinds (corresponding to the cases of 0 to I projections), with proportions $(F_0, \dots, F_I)$. We ask in how many ways we can draw h specific balls with proportions $(f_0, \dots, f_I)$, some of which may be zero. A counting argument leads to the multivariate hypergeometric distribution[7]:

$$\mathrm{p}(\{\underbrace{N_{\dots}=0}_{h f_0 \text{ of them}}\},\ \{\underbrace{N_{\dots}=1}_{h f_1 \text{ of them}}\},\ \dots\ \{\underbrace{N_{\dots}=I}_{h f_I \text{ of them}}\} \mid \boldsymbol{F}, B) = \frac{1}{\binom{h}{h f_0, \dots, h f_I}} \binom{J F_0}{h f_0} \times \dots \times \binom{J F_I}{h f_I} \Big/ \binom{T}{h}. \quad (14)$$

If our question is about a small set of h neurons compared to the total number T, and the proportions f_t are small compared to 1, the formula

[7] Johnson et al. 1996 ch. 39.

above has a useful approximation:

$$\mathrm{p}(\{\underbrace{N_{\dots}=0}_{hf_0 \text{ of them}}\},\ \{\underbrace{N_{\dots}=1}_{hf_1 \text{ of them}}\},\ \dots\ \{\underbrace{N_{\dots}=I}_{hf_I \text{ of them}}\} \mid \boldsymbol{F}, B) \approx \prod_{n=0}^{I} (F_n)^{hf_n}. \qquad (15)$$

The integral appears because we approximate the tuple $\boldsymbol{F}$ as a continuous quantity. It takes values in an $(I+1)$-dimensional space.

We have one such distribution separately for each mouse. But the collection of these distributions depend on the statistics $\{\boldsymbol{F}^{(m)}\}$, which we don't know. We discuss our guess about them next

3.3.3 $\boldsymbol{F}^{(m)}$ *for the mouse population*

So far we have expressed our guess about the projection indicators $\{c_{it}^{(m)}\}$ conditional on the number of projections $\{N_t^{(m)}\}$ and our guess about these numbers conditional on the statistics $\{\boldsymbol{F}^{(m)}\}$. Let's finally consider our guess about these statistics.

We proceed analogously as in the previous two guesses, by assuming that we know the 'super-statistics' $\boldsymbol{S}$ about the statistics $\{\boldsymbol{F}^{(m)}\}$ of all mice. This time, however, we'll also assign a guess for $\boldsymbol{S}$ without relying on higher-level statistics. We also make an analogous symmetry assumption, this time about the mouse population.

The result is analogous to the guess about $\{N_t^{(m)}\}$. This time we use the analogous version of the approximate form (15), since the number of mice is in principle infinite, for the happiness of exterminator companies:

$$\mathrm{p}(\underbrace{\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \dots}_{\text{any number}} \mid \boldsymbol{S}, B) = S_{\boldsymbol{F}^{(1)}} \times S_{\boldsymbol{F}^{(2)}} \times \cdots . \qquad (16)$$

The super-statistics $\boldsymbol{S}$ is unknown to us. Our pre-data guess about it is expressed by a probability density $\mathrm{p}(\boldsymbol{S} \mid B)\,\mathrm{d}\boldsymbol{S}$. It turns out that the exact form of this density is unimportant if we have numerous data, because the data will change its shape to reflect the observed statistics. For the moment we assume it to be constant:

$$\mathrm{p}(\boldsymbol{S} \mid B) = \text{const.} \qquad (17)$$

but later we'll check if changes in this distribution lead to important changes in our final inferences.

3.3.4 *Final form for our pre-data distribution*

The final form of our pre-data distribution is obtained by combining the distributions (13) and (14) for each mouse of interest, together with (16) and (17). And summing over all possible values of $\boldsymbol{S}$, of all $\{\boldsymbol{F}^{(m)}\}$ and of all $\{N_t^{(m)}\}$. We don't write this down, because some of those sums won't be necessary when we calculate the probability of the desired answers given our data.

3.4 Step 4: the probability for the answers given the data

Our final goal is to guess the projection statistics $\boldsymbol{F}^{(m)}$ for the experimentally observed mice $\{m\}$, and the super-statistics $\boldsymbol{S}$ for the whole mouse population under study, given our experimental data and pre-data information:

$$\mathrm{p}(\underbrace{\{\boldsymbol{F}^{(m)}\}}_{\text{observed mice}}, \boldsymbol{S} \mid \text{data}, B). \tag{18}$$

This conditional probability can be calculated in different ways. The most convenient for our problem is Bayes's theorem, because it relates this conditional probability to one with data and statistics inverted:

$$\mathrm{p}(\{\boldsymbol{F}^{(m)}\}, \boldsymbol{S} \mid \text{data}, B) \propto \mathrm{p}(\text{data} \mid \{\boldsymbol{F}^{(m)}\}, \boldsymbol{S}, B) \times \mathrm{p}(\{\boldsymbol{F}^{(m)}\}, \boldsymbol{S} \mid B) \tag{19}$$

The probability distribution on the right is useful because it has the unknown statistics in the conditional, allowing us to use the pre-data probabilities of § 3.3. The product on the right can be found, apart from a normalizing constant, by sampling methods such as Monte Carlo.

We now find the probability of the data, assuming for the moment that all statistics $\boldsymbol{S}$, $\{\boldsymbol{F}^{(m)}\}$, $\{N_t^{(m)}\}$ are known.

The data consist in the two main statements $\alpha^{(m)}$ and $\beta^{(m)}$, discussed in § 3.2.2, for all observed mice $\{m\}$. Let's first focus on one specific mouse and omit '(m)'. We calculate the probability for the data as follows:

$$\mathrm{p}(\alpha, \beta \mid \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) = \mathrm{p}(\alpha \mid \beta, \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) \times \mathrm{p}(\beta \mid \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B). \tag{20}$$

The last factor is easily calculated using formula (13) with $n_0 = I - \hat{I}$ and $n_1 = 0$:

$$\mathrm{p}(\beta \mid \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) \equiv \mathrm{p}(\{\underbrace{c_{\dots,1} = 0}_{I-\hat{I} \text{ of them}}\}, \dots, \{\underbrace{c_{\dots,\hat{T}} = 0}_{I-\hat{I} \text{ of them}}\} \mid \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) =$$

$$\prod_{t=1}^{\hat{T}} \binom{N_t}{0} \binom{I-N_t}{I-\hat{I}} \Big/ \binom{I}{I-\hat{I}} \equiv \prod_{t=1}^{\hat{T}} \binom{\hat{I}}{N_t} \Big/ \binom{I}{N_t}. \quad (21)$$

This results makes sense: for each column t, the denominator $\binom{I}{N_t}$ counts the total number of ways in which N_t black entries can be distributed among $\hat{I}$ places. The numerator $\binom{\hat{I}}{N_t}$ counts the number of ways in which the N_t black entries can be limited to the first $\hat{I}$ places, leaving the remaining $I - \hat{I}$ blank, as requested by our data β.

Now to $\mathrm{p}(\alpha \mid \beta, \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B)$. By the Boolean expression of α, eq. (12),

$$\begin{aligned}
\mathrm{p}(\alpha \mid \beta, \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) &\equiv \mathrm{p}\Big\{\neg \bigvee_{i=1}^{\hat{I}} \Big[\bigwedge_{t=1}^{\hat{T}} (c_{it} = 0)\Big] \mid \dots\Big\} = \\
&1 - \mathrm{p}\Big\{\bigvee_{i=1}^{\hat{I}} \Big[\bigwedge_{t=1}^{\hat{T}} (c_{it} = 0)\Big] \mid \dots\Big\} = \\
&1 - \sum_{i=1}^{\hat{I}} \mathrm{p}\Big[\bigwedge_{t=1}^{\hat{T}} (c_{it} = 0) \mid \dots\Big] \\
&\quad + \sum_{i'>i''}^{i',i''} \mathrm{p}\Big[\bigwedge_{t=1}^{\hat{T}} (c_{i't} = 0) \;\; \bigwedge_{t=1}^{\hat{T}} (c_{i''t} = 0) \mid \dots\Big] \\
&\quad - \cdots \text{ [terms with three and more } i\text{]}. \quad (22)
\end{aligned}$$

Where we have indicated the conditional with $\dots$ for brevity. For the first equality we have used the rule for probability of the negation of a statement. For the second equality we have used the sum rule of probability for *non*-mutually exclusive statements. In fact, the expression $\bigwedge_{t=1}^{\hat{T}} (c_{it} = 0)$ says that the ith row is empty, but this doesn't exclude that other rows be empty as well.

Each summand in the last expression is calculated using formula (13). Consider the generic term with r different is:

$$(-1)^r \mathrm{p}\Big[\underbrace{\textstyle\bigwedge_{t=1}^{\hat{T}}(c_{\dots t}=0) \;\cdots\; \bigwedge_{t=1}^{\hat{T}}(c_{\dots t}=0)}_{r \text{ of these}} \mid \dots\Big] \equiv$$

$$(-1)^r \prod_{t=1}^{\hat{T}} \mathrm{p}\Big[\underbrace{(c_{\dots t}=0), \dots, (c_{\dots t}=0)}_{r \text{ of them}} \mid N_t, \dots\Big] =$$

$$(-1)^r \prod_{t=1}^{\hat{T}} \left[\binom{N_t}{0} \binom{\hat{I}-N_t}{r} \Big/ \binom{\hat{I}}{r} \right] \equiv$$

$$(-1)^r \prod_{t=1}^{\hat{T}} \left[\binom{\hat{I}-r}{N_t} \Big/ \binom{\hat{I}}{N_t} \right]. \tag{23}$$

The first equivalence is just a regrouping of the terms with the same ts together. The second equality is just the application of the pre-data distribution (13) to this case, with $n_0 = r$ and $n_1 = 0$. Note that we have $\hat{I}$ instead of I because we are limited to the first $\hat{I}$ row, the rest being blank as indicated by β in the conditional. The third equivalence is just a combinatorial rearrangement. explain why the result makes sense

Looking again at formula (22), we see that the generic term for r rows above appears for all possible combinations of r different rows out of $\hat{I}$. There are $\binom{\hat{I}}{r}$ such terms, all equal. Also, we note that the generic term (23) for $r = 0$ equals 1. We can therefore rewrite the probability (22) for α as

$$\mathrm{p}(\alpha \mid \beta, \{N_t\}, \boldsymbol{F}, \boldsymbol{S}, B) = \sum_{r=0}^{\hat{I}} (-1)^r \binom{\hat{I}}{r} \prod_{t=1}^{\hat{T}} \left[\binom{\hat{I}-r}{N_t} \Big/ \binom{\hat{I}}{N_t} \right]. \tag{24}$$

Finally, combining this with the probability (21) for β, simplifying the combinatorial terms, and considering all observed mice m we obtain

$$\mathrm{p}(\text{data} \mid \{N_t^{(m)}\}, \{\boldsymbol{F}^{(m)}\}, \boldsymbol{S}, B) =$$

$$\prod_m \sum_{r=0}^{\hat{I}^{(m)}} (-1)^r \binom{\hat{I}^{(m)}}{r} \prod_{t=1}^{\hat{T}^{(m)}} \left[\binom{\hat{I}^{(m)}-r}{N_t^{(m)}} \Big/ \binom{I^{(m)}}{N_t^{(m)}} \right]. \tag{25}$$

The probability above is conditional on $\boldsymbol{S}$, $\{\boldsymbol{F}^{(m)}\}$, $\{N_t^{(m)}\}$; but the latter don't really interest us. We can eliminate them from the conditional using the marginalization rule

$$\mathrm{p}(\mathrm{data}\,|\,\{\boldsymbol{F}^{(m)}\},\boldsymbol{S},B) =$$

$$\sum_{\underbrace{N_t^{(m)}=0}_{\text{for all } m \text{ and } t}}^{I^{(m)}} \cdots \sum \mathrm{p}(\mathrm{data}\,|\,\{N_t^{(m)}\},\{\boldsymbol{F}^{(m)}\},\boldsymbol{S},B)\times \mathrm{p}(\{N_t^{(m)}\}\,|\,\{\boldsymbol{F}^{(m)}\},\boldsymbol{S},B). \quad (26)$$

The first factor in the sum is the probability (25); the second is the pre-data probability (14). Substituting these and rearranging the terms having the same m we find

$$\mathrm{p}(\mathrm{data}\,|\,\{\boldsymbol{F}^{(m)}\},\boldsymbol{S},B) =$$

$$\prod_m \sum_{N_1^{(m)}} \cdots \sum_{N_{T^{(m)}}^{(m)}} \prod_{t=1}^{T^{(m)}} \left[F^{(m)}_{N_t^{(m)}}\right] \sum_{r=0}^{\hat{I}^{(m)}} (-1)^r \binom{\hat{I}^{(m)}}{r} \prod_{t=1}^{\hat{T}^{(m)}} \left[\binom{\hat{I}^{(m)}-r}{N_t^{(m)}} \Big/ \binom{I^{(m)}}{N_t^{(m)}}\right] =$$

$$\prod_m \sum_{r=0}^{\hat{I}^{(m)}} (-1)^r \binom{\hat{I}^{(m)}}{r} \left[\sum_{N=0}^{\hat{I}^{(m)}-r} F^{(m)}_N \binom{\hat{I}^{(m)}-r}{N} \Big/ \binom{I^{(m)}}{N}\right]^{T^{(m)}}. \quad (27)$$

We can finally use Bayes's theorem (19) together with our pre-data probabilities (16) and (17) for the $\{\boldsymbol{F}^{(m)}\}$ and $\boldsymbol{S}$, to obtain the probabilities we're looking for:

$$\mathrm{p}(\underbrace{\{\boldsymbol{F}^{(m)}\}}_{\text{observed mice}},\boldsymbol{S}\,|\,\mathrm{data},B) \propto$$

$$\prod_m S_{\boldsymbol{F}^{(m)}} \sum_{r=0}^{\hat{I}^{(m)}} (-1)^r \binom{\hat{I}^{(m)}}{r} \left[\sum_{N=0}^{\hat{I}^{(m)}-r} F^{(m)}_N \binom{\hat{I}^{(m)}-r}{N} \Big/ \binom{I^{(m)}}{N}\right]^{T^{(m)}}$$

$$\text{with data} = \{\hat{I}^{(m)},\hat{T}^{(m)},I^{(m)},T^{(m)}\}. \quad (28)$$

This probability will be found by Monte Carlo sampling.

4 Tying up loose ends

In the probability (28) we have defined our data to be the number of marked input and target neurons in each mouse, but also the *total* number of neurons in the input area and of the specified target class (and target region). We actually don't have the latter numbers. In principle it would be possible to guess them too, but the calculations would become much heavier. Instead we rely on the fact that the marked neurons are far fewer than the total, $\hat{I}^{(m)} \ll I^{(m)}$, $\hat{T}^{(m)} \ll T^{(m)}$, fixing the totals to a high plausible number (100 000). Possibly we'll also consider treating these numbers as infinities.

5 On hierarchic modelling

Bibliography

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

Barwise, J., Etchemendy, J. (2003): *Language, Proof and Logic*. (CSLI, Stanford). Written in collaboration with Gerard Allwein, Dave Barker-Plummer, Albert Liu. First publ. 1990.

Damien, P., Dellaportas, P., Polson, N. G., Stephens, D. A., eds. (2013): *Bayesian Theory and Applications*. (Oxford University Press, Oxford).

Dawid, A. P. (2013): *Exchangeability and its ramifications*. In: Damien, Dellaportas, Polson, Stephens (2013), ch. 2, 19–29.

de Finetti, B. (1930): *Funzione caratteristica di un fenomeno aleatorio*. Atti Accad. Lincei: Sc. Fis. Mat. Nat. **IV**^{5}, 86–133. http://www.brunodefinetti.it/Opere.htm.

Diaconis, P. (1977): *Finite forms of de Finetti's theorem on exchangeability*. Synthese **36**^{2}, 271–281. http://statweb.stanford.edu/~cgates/PERSI/year.html.

Diaconis, P., Freedman, D. (1980): *Finite exchangeable sequences*. Ann. Prob. **8**^{4}, 745–764.

Good, I. J. (1969): *A subjective evaluation of Bode's law and an 'objective' test for approximate numerical rationality*. J. Am. Stat. Assoc. **64**^{325}, 23–49. Partly repr. in Good (1983) ch. 13.

— (1983): *Good Thinking: The Foundations of Probability and Its Applications*. (University of Minnesota Press, Minneapolis, USA).

Hailperin, T. (1996): *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*. (Associated University Presses, London).

— (2011): *Logic with a Probability Semantics: Including Solutions to Some Philosophical Problems*. (Lehigh University Press, Plymouth, UK).

Heath, D., Sudderth, W. (1976): *De Finetti's theorem on exchangeable variables*. American Statistician **30**^{4}, 188–189.

Hewitt, E., Savage, L. J. (1955): *Symmetric measures on Cartesian products*. Trans. Am. Math. Soc. **80**^{2}, 470–501.

Jaynes, E. T. (2003): *Probability Theory: The Logic of Science*. (Cambridge University Press, Cambridge). Ed. by G. Larry Bretthorst. First publ. 1994. https://archive.org/details/XQUHIUXHIQUHIQXUIHX2, http://www-biba.inrialpes.fr/Jaynes/prob.html.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1996): *Discrete Multivariate Distributions*. (Wiley, New York). First publ. 1969 in chapter form.

Johnson, W. E. (1924): *Logic. Part III: The Logical Foundations of Science*. (Cambridge University Press, Cambridge). https://archive.org/details/logic03johnn.

Sox, H. C., Higgins, M. C., Owens, D. K. (2013): *Medical Decision Making*, 2nd ed. (Wiley, New York). First publ. 1988.