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 21 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.

This document is designed for screen reading and two-up printing on A4 or Letter paper

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.

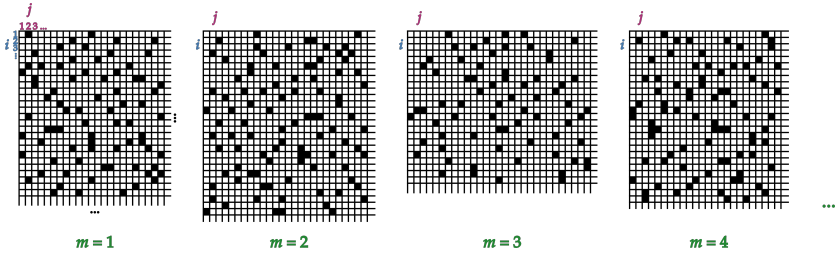

Figure 1 Imaginary layout of the connectivity from input to target region for all mice

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

$\boldsymbol{H}^{(m)} \coloneqq \big(H_0^{(m)}, H_1^{(m)}, H_2^{(m)}, \dots, H_{I^{(m)}}^{(m)}\big)$ 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

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

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

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

similarly to the expression for the individual statistics (32). 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

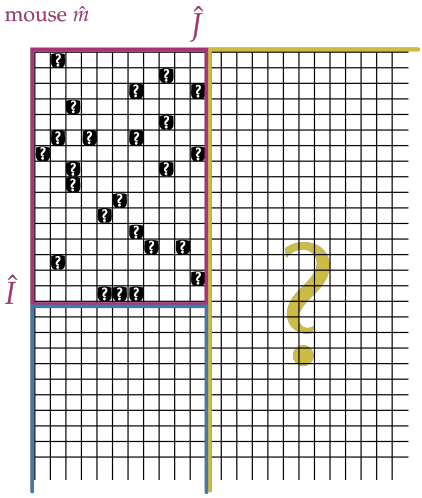

Figure 2 Imaginary layout of the connectivity data for an observed mouse

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 (27), 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 (26) for each mouse of interest, together with (28) and (30). 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

$$\mathrm{p}(\text{data} \mid \{N_t^{(m)}\}, \{\boldsymbol{F}^{(m)}\}, \boldsymbol{S}, B) = \prod_m \sum_{r=0}^{\hat{I}^{(m)}} (-1)^r \binom{\hat{I}^{(m)}}{r} \prod_{t=1}^{\hat{T}^{(m)}} \left[\binom{\hat{I}^{(m)}-r}{N_t^{(m)}}\Big/\binom{I^{(m)}}{N_t^{(m)}}\right]. \tag{25}$$

[from here on: old text]

In other words, the specific index j of a neuron shouldn't give us any information about the number of projections N_j to it. This is called an *exchangeability* assumption. To find this probability it's easier to first form the histogram $\boldsymbol{h} = (h_0, \dots, h_I)$ of these values. We do this by first grouping together all N_j that have the same value. For example, if the first, third, and tenth neurons have five projections, then $N_1 = N_3 = N_{10} = 5$ and we group them together. There are $I + 1$ groups: the group with no projection, that is, with all $N_j = 0$, up to the group with I projections, that is, with all $N_j = I$. Count how many N_j are in the group with n projections: that's h_n.

The histogram $\boldsymbol{h}$ is useful because *all sets of values $N_1, \dots, N_{J'}$ leading to the same histogram have the same probability*. This is a consequence of exchangeability: it is as likely that first and second neurons have 4 projections each, and the third has 6 projection, as that the first and the third have 4 projections and the second has 6 instead, because the index j doesn't matter. But either of these cases leads to an histogram with $h_4 = 2$ and $h_6 = 1$.

De Finetti's theorem say that the probability of observing a set of projections $N_1, \dots, N_{J'}$ having the histogram $\boldsymbol{h}$ is

$$\mathrm{p}(\underbrace{N_1, \dots, N_{J'}}_{\text{histogram } \boldsymbol{h}} \mid B) = \frac{1}{J'} \sum_{\boldsymbol{H}} \mathrm{p}(\boldsymbol{H} \mid B) \, \frac{\binom{H_0}{h_0} \times \dots \times \binom{H_n}{h_n}}{\binom{J}{J'}}, \tag{26}$$

where $\boldsymbol{H}$ is the histogram for the full set of J target neurons, previously discussed. The sum ranges over all possible such histograms, and $\mathrm{p}(\boldsymbol{H} \mid B)$ is our pre-data probability distribution for them. The fraction is a multivariate hypergeometric distribution[8] characteristic of 'sampling without replacement'. The formula above corresponds to first guessing the histogram $\boldsymbol{H}$ for all target neurons, and then 'sampling' J' out of them.

If our question is about a small set of J' neurons compared to the total number J, and the values N_j are small compared to their theoretical maximum I, the formula above has a useful approximation. Denote

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

by $\boldsymbol{F} \coloneqq \boldsymbol{H}/J$ the normalized histogram, corresponding to the relative frequency distribution of projections. Then

$$\mathrm{p}(\underbrace{N_1, \dots, N_{J'}}_{\text{histogram } \boldsymbol{h}} \mid B) \approx \int \mathrm{d}\boldsymbol{F}\ \mathrm{p}(\boldsymbol{F} \mid B)\ F_{N_1} \times \cdots \times F_{N_{J'}} \equiv \int \mathrm{d}\boldsymbol{F}\ \mathrm{p}(\boldsymbol{F} \mid B)\ \prod_{n=0}^{I} {F_n}^{h_n}. \quad (27)$$

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

We still have to specify the pre-data distribution $\mathrm{p}(\boldsymbol{H} \mid B)$ or $\mathrm{p}(\boldsymbol{F} \mid B)$ for the full histogram $\boldsymbol{H}$ or its normalized version $\boldsymbol{F}$. We do this in the next section, where we consider our inferences about all mice.

3.4.1 *All mice: exchangeability*

We have now the pre-data distribution for the numbers of projections, as well as their histogram, in a single mouse: $\mathrm{p}(\{N_j\} \mid B)$. How to extend this to the projections in all mice, $\mathrm{p}\big(\{N_j^{(1)}\}, \{N_j^{(2)}\}, \dots \mid B\big)$?

Also in this case we invoke symmetry, this time for the frequency distribution of projections $\boldsymbol{F}^{(m)}$ that a mouse can have. Our assumption is this: *there isn't any reason for some mouse m to have a specific distribution of projections $\boldsymbol{F}$ more than another mouse m'*. In other words, we consider all mice under study, past and future, to be similar regarding their distributions of projections. This makes sense because we want to make some generalized statements about mice within a specific group, for example with specific age, gender, or genetic modifications. Our experimental observations always exclude any mice that don't have the requirements defining the group under study; hence the assumption above is satisfied by design. add note about unknown biases

This assumption is again one of exchangeability, and it leads to a result similar to that of the previous section. The pre-data distribution for the histograms $\{\boldsymbol{H}^{(m)}\}$ of several mice is given by

$$\mathrm{p}(\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \dots \mid B) = \int \mathrm{d}\boldsymbol{S}\ \mathrm{p}(\boldsymbol{S} \mid B)\ S_{\boldsymbol{F}^{(1)}} \times S_{\boldsymbol{F}^{(2)}} \times \cdots\ . \quad (28)$$

Here $\boldsymbol{S}$ is the distribution of all normalized histograms $\boldsymbol{F}$. That is, we imagine to check every single mouse, and for each we write down its

histogram of projections $\boldsymbol{F}$; 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. Thus $S_{\boldsymbol{F}}$ is the proportion of mice that have the same histogram $\boldsymbol{F}$; it's a sort of 'super-statistics'. Mathematically it's expressed as

$$S_f \coloneqq \frac{1}{M} \sum_m \delta\bigl(\boldsymbol{F}^{(m)} = f\bigr), \tag{29}$$

similarly to the expression for the individual statistics (32). 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 $\boldsymbol{F}$ is a tuple with $I+1$ elements, the distribution $\boldsymbol{S}$ is over an $(I+1)$-dimensional space.

The distribution $\boldsymbol{S}$ is unknown to us, and the probability density $\mathrm{p}(\boldsymbol{S} \mid B)$ expresses our pre-data uncertainty about it. We still have to specify this density. It turns out that its exact form 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:

$$\mathrm{p}(\boldsymbol{S} \mid B) = \text{const.} \tag{30}$$

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

We can finally combine the pre-data distributions (27), (28), and (30) to obtain the pre-data distribution for all combinations of atomic statements:

$$\begin{aligned}\mathrm{p}(\underbrace{N_1^{(1)}, N_2^{(1)}, \dots,}_{\text{histogram } \boldsymbol{h}^{(1)}} \underbrace{N_1^{(2)}, N_2^{(2)}, \dots,}_{\text{histogram } \boldsymbol{h}^{(2)}} \dots \mid B) \propto \\ \int \mathrm{d}\boldsymbol{S} \prod_m \biggl[\int \mathrm{d}\boldsymbol{F}^{(m)}\, S_{\boldsymbol{F}^{(m)}} \prod_j F^{(m)}_{N_j} \biggr] \equiv \\ \int \mathrm{d}\boldsymbol{S} \prod_m \biggl[\int \mathrm{d}\boldsymbol{F}^{(m)}\, S_{\boldsymbol{F}^{(m)}} \prod_{n=0}^{I^{(m)}} \bigl(F^{(m)}_n\bigr)^{h^{(m)}_n} \biggr].\end{aligned} \tag{31}$$

Note what the formula above is doing: we have one expression within brackets for each mouse m; this expression gives a probability to the

projections N_j of any target neurons we please, based on the full statistics $\boldsymbol{F}^{(m)}$ for that mouse. Such full statistics for all mice are in turn given a probability based on a 'super-statistics' $\boldsymbol{S}$ for the full mouse population.

These statements won't be our atomic statements, however: it's more profitable to use slightly less detailed ones. We're interested, in fact, in the total number of neurons in an input region that project to each target neuron j, but not on their individual identities. Let $N_j^{(m)}$ denote this number. In fig. 1 it's the total number of black squares in column j of grid m. Its possible values are between 0 and $I^{(m)}$.

If we could experimentally count the projections and find $N_j^{(m)}$ for all $T^{(m)}$ target neurons of mouse m, we could then report several statistics; for example:

- the average number of projections to the target neurons in that mouse: $\overline{N}^{(m)} \coloneqq \frac{1}{T^{(m)}} \sum_j N_j^{(m)}$;

- the variance from that average number, $\frac{1}{T^{(m)}} \sum_j \bigl(N_j^{(m)} - \overline{N}^{(m)}\bigr)^2$, or any other measure of variation;

- the full histogram $\boldsymbol{H}^{(m)} \coloneqq \bigl(H_0^{(m)}, H_1^{(m)}, H_2^{(m)}, \dots, H_{I_m}^{(m)}\bigr)$ of the number of projections for that mouse. That is: we count how many target neurons receive no projection from the input region, that's $H_0^{(m)}$; we count how many receive projections from only one neuron, that's $H_1^{(m)}$; and so on. Mathematically this is written

$$H_n^{(m)} \coloneqq \sum_j \delta\bigl(N_j^{(m)} = n\bigr) \tag{32}$$

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

If we had these statistics and histograms $\boldsymbol{H}^{(m)}$ for each mouse m, we could then ask about the average histogram of projections $\bar{\boldsymbol{F}} \coloneqq \frac{1}{M} \sum_m \boldsymbol{F}^{(m)}$ across the whole mouse population, where M is the total number of mice. We could also ask about the variance from such an average histogram. These histograms and their average would be the answers to the 'individual' and 'collective' questions we asked in § 2.

So, *our atomic statements will be about the values of the quantities* $\{N_j^{(m)}\}$, for each target neuron j of each mouse m. If we give a probability

distribution for these quantities,

$$\mathrm{p}(\{N_j^{(m)}\} \mid B) \equiv \mathrm{p}\big(N_1^{(1)}, N_2^{(1)}, \dots, N_{T^{(1)}}^{(1)}, N_1^{(2)}, N_2^{(2)}, \dots, N_{T^{(2)}}^{(2)}, \dots \mid B\big), \tag{33}$$

then we can calculate from it the probabilities (3) for the answers to our main questions. We do this in the following section.

In the next step we express our experimental data in terms of the atomic statements about the projection numbers $\{N_j^{(m)}\}$. This will allow us to use the probability (31) in reverse: from the $\{N_j^{(m)}\}$ that make up the data we will make guesses about the single-mouse statistics $\boldsymbol{F}^{(m)}$ and the mouse-population statistics $\boldsymbol{S}$.

3.5 Step 3: data and answers in terms of atomic statements

Let's start with the answers to our main questions of p. 2. As discussed in § 3.1, these answers are given by the statistics $\boldsymbol{F}^{(m)}$ and $\boldsymbol{S}$, already expressed in terms of the projection numbers $\{N_j^{(m)}\}$ via formulae (32) and (29). And these statistics already appear in the pre-data probabilities (31), we don't need to analyse them further.

In order to assign the initial joint probability distribution $\mathrm{p}[(c_{itm}) \mid B]$, for all values of the connections (c_{mij}), we explore the symmetries of our problem. Symmetries translate into equalities of some values of the joint distribution, and therefore restrict its possible values and mathematical form.

Let's give an example corresponding to the three-level hierarchy **??** of § 3.0. If we consider all input neurons (i) to a specific target neuron j in a specific mouse m as similar, then the joint probability should stay the same under exchanges of the indices i for which the connections

have the same values. For example:

$$\begin{aligned}
&\mathrm{p}(c_{2\,jm}=0, c_{5\,jm}=1, c_{6\,jm}=1, c_{9\,jm}=0 \mid B) = \\
&\qquad \mathrm{p}(\underbrace{c_{9\,jm}=0, c_{5\,jm}=1, c_{6\,jm}=1, c_{2\,jm}=0}_{2\leftrightarrow 9} \mid B) = \\
&\qquad \mathrm{p}(\underbrace{c_{2\,jm}=0, c_{6\,jm}=1, c_{5\,jm}=1, c_{9\,jm}=0}_{5\leftrightarrow 6} \mid B) = \\
&\qquad\qquad \mathrm{p}(\underbrace{c_{9\,jm}=0, c_{6\,jm}=1, c_{5\,jm}=1, c_{2\,jm}=0}_{2\leftrightarrow 9,\quad 5\leftrightarrow 6} \mid B). \qquad (34)
\end{aligned}$$

We can't exchange $i = 2$ and $i = 5$ because their connections have different values. Because of this symmetry in i the joint probability must have a specific form, by de Finetti's representation theorem:[9]

$$\mathrm{p}[(c_{ijm}) \mid B] = \int \prod_{jm} \Big[\mathrm{d}f_{jm}\; f_{jm}^{\sum_i c_{ijm}}\, (1-f_{jm})^{\sum_i (1-c_{ijm})} \Big]\; \mathrm{p}[(f_{jm}) \mid B]. \qquad (35)$$

In this expression, f_{jm} is the total number of neurons projecting to target neuron j in mouse m, and what remains to be assigned is the probability distribution $\mathrm{p}[(f_{jm}) \mid B]$ over such numbers for all j and m. It may look like a complicated expression, but it has considerably reduced the possible assignments of joint probabilities.

The second level of the hierarchy of this example was to consider a similarity of connectivity statistics – which are the quantities f_{jo} – among all target neurons in each specific mouse. This means that the joint probability $\mathrm{p}[(f_{jm}) \mid B]$ in the formula above is symmetric in the index j for each fixed m. This leads, again by de Finetti's theorem, to a new expression similar to the one above, with new m-dependent statistics as parameters and a joint distribution for htem. The final level in the hierarchy says that this distribution is itself symmetric in the index m, thus leading to a final integral expression in terms of a single parameter only with an associated probability distribution.

The difficulty of the three-level formulation just sketched is mathematical: the integrals involved are over more complex spaces which are numerically more expensive to handle.

[9] De Finetti 1930; Hewitt et al. 1955; Heath et al. 1976; Diaconis 1977; Diaconis et al. 1980; Dawid 2013.

As a first, simpler formulation we consider instead the more unrealistic but powerful symmetry of example **??**: the joint distribution $\mathrm{p}[(c_{ijm}) \mid B]$ stays the same under exchanges of *any* indices (i, j, m) for which the connections have the same values. For example,

$$\begin{aligned}\mathrm{p}(c_{237}=0, c_{552}=1, c_{914}=0 \mid B) = \mathrm{p}(c_{937}=0, c_{552}=1, c_{214}=0 \mid B) = \\ \mathrm{p}(c_{217}=0, c_{552}=1, c_{934}=0 \mid B) = \mathrm{p}(c_{234}=0, c_{552}=1, c_{917}=0 \mid B) = \\ \mathrm{p}(c_{917}=0, c_{552}=1, c_{234}=0 \mid B) = \cdots \quad \text{and so on.}\end{aligned} \tag{36}$$

We can only exchange indices with the same connection values, but we have full freedom otherwise.

Again by de Finetti's theorem, this symmetry forces our joint probability distribution to have this form:

$$\begin{aligned}\mathrm{p}[(c_{ijm}) \mid B] &= \int \mathrm{d}q \; q^{\sum_i c_{ijm}} \, (1-q)^{\sum_i (1-c_{ijm})} \, \mathrm{p}(q \mid B) \\ &\equiv \int \mathrm{d}q \; q^{Rf} \, (1-q)^{R(1-f)} \, \mathrm{p}(q \mid B)\end{aligned} \tag{37}$$

The symmetries briefly discussed previously

Let's analyse two approximate symmetries first. Consider a specific mouse m.

For each neuron i in the input region, we can approximately judge as equally likely that it will connect to any of the neurons j in the target region. This means that our initial probability distribution should remain the same if we swap some values of the index j, with m and i fixed. For example, the probability that neuron #3 in the input region projects to neuron #2 in the target region and does not project to neuron #5 should be the same if we exchange 2 and 5 in that statement. And similarly for combinations of such statements:

$$\mathrm{p}(c_{m\,32}=1, c_{m\,35}=0 \mid B) = \mathrm{p}(c_{m\,32}=0, c_{m\,35}=1 \mid B). \tag{38}$$

This symmetry is an approximation: if the indices i and j bear some information about the neurons' locations, our probability should reflect the possibility that nearby neurons may connect more easily, for example.

We have an analogous approximate symmetry under exchanges of the input-neuron index i for fixed target neuron j.

These two approximate symmetries together are important because they greatly restrict the possible values of our initial probability distribution $\mathrm{p}[(c_{mij}) \mid B]$. The result they lead to[10] is mathematically difficult for me, however. To further simplify the problem I make an even rougher approximation: that the symmetries above hold not just for i and j separately, but for any pairs of them. This assumption is even more unrealistic than the previous two, because it gives the same probability to the situation where each input neuron is connected with only one, distinct, target neuron, as to the situation where only one input neuron is connected to all target neurons, and all other input neurons are unconnected.

With this approximation the joint probability distribution for a given mouse m assumes the following form, by de Finetti's representation theorem:[11]. Take a set of k_m input neurons and l_m target neurons, and consider all possible connections c_{mij} between them, with $i = 1, \dots, k_m$ and $j = 1, \dots, l_m$. Then

$$\begin{aligned} \mathrm{p}[(c_{mij}) \mid m, B] &= \int_0^1 \mathrm{d}\nu_m \; \mathrm{p}(\nu_m \mid B) \; \prod_{ij} \left[{\nu_m}^{c_{mij}} \, (1-\nu_m)^{1-c_{mij}} \right] \\ &\equiv \int_0^1 \mathrm{d}\nu_m \; \mathrm{p}(\nu_m \mid B) \; {\nu_m}^{k_m l_m f_m} \, (1-\nu_m)^{k_m l_m (1-f_m)} . \end{aligned} \tag{39}$$

Here $k_m l_m f_m = \sum_{ij} c_{mij}$ is the total number of connections between the k_m input neurons and the l_m target neurons, and $0 \leqslant \nu_m \leqslant 1$ is the population-averaged number of connections from the *whole* input region to the *whole* target region. The probability distribution $\mathrm{p}(\nu_m \mid I)$ expresses our beliefs about the total number of connections from input to target region in mouse m. The formula above is valid only if k_m and l_m are small (say, one tenth at most) compared to their region sizes.

Let's see how our question and data can be expressed in term of these statements and the quantities (c_{mij}).

We imagine to check every neuron i in the target region of every mouse m. For each such neuron we count how many neurons from the input region project to it. Let's call this number N_{mi}. From the point of view of fig. 1 we're considering every column from every grid, in turn,

[10] Hoover 1979; Aldous 1981; Diaconis et al. 1981. [11] De Finetti 1930; Hewitt et al. 1955; Heath et al. 1976; Diaconis 1977; Diaconis et al. 1980; Dawid 2013.

and counting how many black entries it has. If we were able to perform these observations, at the end we could build a histogram of the numbers (N_{mi}), showing the proportion of neurons in the target region that have 0 inputs from the input region, or 1 input, 2 inputs, n inputs, and so on. These values would be the bins of the histogram. The number of bins is large – as much as the largest number of neurons that could make the input region of a mouse. Denote by F_n the proportion (between 0 and 1) of neurons that receive n inputs. This proportion is calculated from all numbers (N_{mi}) by

$$F_n = \frac{\sum_{mi} \delta(N_{mi} = n)}{\sum_m L_m} \equiv \frac{\text{number of } N_{mi}\text{s equal to } n}{\text{total number of neurons in target regions}}. \quad (40)$$

The whole histogram is the set of proportions $\boldsymbol{F} \coloneqq (F_0, F_1, \dots)$.

This histogram can give one definite meaning to the question 'how many inputs from the input region does a neuron in the target region have on average?': we are asking 'what the mean of the histogram?', that is,

$$\text{average number of inputs} = \sum_n n\, F_n \equiv \frac{\sum_{mi} N_{mi}}{\sum_m L_m}. \quad (41)$$

But the histogram would tell us even more, for example the percentage of target neurons that receive between 100 and 200 inputs from the input region, obtained by summing $F_{100} + \dots + F_{200}$.

It's good to keep in mind that our question could also be approached in slightly different ways, leading to different secondary questions. For example, we could construct a histogram for the number of inputs for each mouse individually, thus obtaining several $\boldsymbol{F}^{(m)} = \left(F_0^m, F_1^m, \dots\right)$, one for each m. Then we could ask: 'what is the average connectivity histogram for a mouse?', that is,

$$\frac{1}{\text{number of mice}} \sum_m \boldsymbol{F}^{(m)} \quad (42)$$

and also ask how much such histogram can vary from mouse to mouse.

These alternative questions are important in view of possible approximations in our assumptions and maths.

For example, if we assume that all target neurons from all mice can be somehow pooled together – which makes the maths easier – then we are forsaking the possibility of analysing the individual connectivity histograms.

Although we'll never be able to construct such a histogram, we can nevertheless guess, from our observations in a small set of mice, what kind of shape and numerical values it could have. So our goal is to determine the probability

$$\mathrm{p}(\text{histogram} \mid \text{data}, B) \tag{43}$$

for each possible histogram, given the data collected in our observations and any other information B we have, including our present knowledge of brain biology.

These probability above can be found using the rule for conditional probabilities:

$$\mathrm{p}(\text{histogram} \mid \text{data}, B) \propto \mathrm{p}(\text{data} \mid \text{histogram}, B) \times \mathrm{p}(\text{histogram} \mid B). \tag{44}$$

To calculate these we need to determine our initial probability distribution $\mathrm{p}(\text{histogram} \mid B)$ for all possible histograms. In fact we need to determine the initial probability of a couple more things, for example the total number of neurons in the input region K_m, for each mouse m. All other probabilities are calculated from this initial probability distribution.

In the next section we discuss and analyse several alternative assumptions that we can use to build the distribution $\mathrm{p}(\text{histogram} \mid B) \equiv \mathrm{p}(\boldsymbol{F} \mid B)$. Some of these assumptions are approximate or unrealistic, yet unfortunately necessary for making the problem mathematically tractable. After discussing the alternatives we'll choose specific ones to arrive at formulae that we can use with the data. The discussion of the alternatives is useful because we can step back and try to change the less realistic assumptions if we see that our formulae don't work well.

4 Initial probability distribution

The assumptions we use to build our initial probability distribution $\mathrm{p}(\boldsymbol{F} \mid B)$ are about the symmetries of our problem.

The first symmetry is very intuitive and compelling. Take the distribution of numbers of inputs for a given mouse m: (N_{m1}, N_{m2}) ***

In order to formulate the initial probability distribution $\mathrm{p}[(c_{mij}) \mid B]$ we explore the symmetries of our problem.

Let's analyse two approximate symmetries first. Consider a specific mouse m.

the answers to our question Q2 and the statement of our experimental observations are built up from the specific statements (48). For example, the statement *In mouse # 3, neuron # 1 in the target region receives input from 2 neurons in the input region* is equivalent to the composite statement *In mouse # 3, neuron # 1 in the target region receives input from neurons # 1 and # 2, or # 1 and # 3, or # 1 and # 4, or … in the input region*, which we can express as

$$(C_{311} \text{ and } C_{312}) \text{ or } (C_{311} \text{ and } C_{313}) \text{ or } \cdots \\ \text{or } (C_{312} \text{ and } C_{313}) \text{ or } (C_{312} \text{ and } C_{314}) \text{ or } \cdots \tag{45}$$

Let's denote such kind of statement with

$$C_{mij} \coloneqq \textit{In mouse } m\textit{, neuron } i \textit{ in input region projects to neuron } j \textit{ in target region} \tag{46}$$

and let's introduce the quantity

$$c_{mij} = \begin{cases} 1 & \text{if } C_{mij} \text{ is true,} \\ 0 & \text{if } C_{mij} \text{ is false.} \end{cases} \tag{47}$$

We can rephrase it in two different ways; it's the second that probably interests us:

Q1. If we examine a new neuron in region A in a new mouse, how many neurons in region B will project to it? The possible answers to this question are: 0, 1, 2, and so on.

Q2. If we examine all neurons in region A in a very large sample of mice, and for each neuron we count how many neurons from region B project to it, then what are the frequencies with which we'll observe 0 connections, 1 connection, 2 connections, and so on? One possible answer could be this distribution:

$$0\text{: } 0.1\%, \quad 1\text{: } 0.3\%, \quad \ldots, \quad 1\,000\text{: } 24\%, \quad \ldots$$

another answer could be this distribution:

$$0\text{: } 0.02\%, \quad 1\text{: } 0.07\%, \quad \ldots, \quad 1\,000\text{: } 31\%, \quad \ldots$$

and so on, for all possible distributions of frequencies.

For the first question we must give a distribution of probability over all possible answers 0, 1, and so on. For the second the probability is distributed over all possible frequency distributions.

The answers to these two questions and their probabilities are connected; in particular, from the second we can obtain the first. The second is more informative, because it gives us also a range of variability.

Our analysis will also produce probabilities for other another quantity: how many neurons are there in region B, in the next mouse we observe?

We must consider the set of mice that have been observed and those that will be observed in the future. Let's label each mouse with an index m. For our inferences to be valid, these mice must be judged 'similar'.

Now that we've labelled the mice and their neurons in the two regions, we can make several concrete statements, such as

> *In mouse # 64, the neuron # 931 in the input region projects to the neuron # 1 488 in the target region.*

which can be either true or false. Let's denote such kind of statement with

$$C_{mij} \coloneqq \textit{In mouse } m\textit{, neuron } i \textit{ in input region projects to neuron } j \textit{ in target region} \tag{48}$$

and let's introduce the quantity

$$c_{mij} = \begin{cases} 1 & \text{if } C_{mij} \text{ is true,} \\ 0 & \text{if } C_{mij} \text{ is false.} \end{cases} \tag{49}$$

Obviously for most of these statements we'll never be able to ascertain whether they're true or false. But the answers to our question Q2 and the statement of our experimental observations are built up from the specific statements (48). For example, the statement *In mouse # 3, neuron # 1 in the target region receives input from 2 neurons in the input region* is equivalent to the composite statement *In mouse # 3, neuron # 1 in the target region receives input from neurons # 1 and # 2, or # 1 and # 3, or # 1 and # 4, or … in the input region*, which we can express as

$$(C_{311} \text{ and } C_{312}) \text{ or } (C_{311} \text{ and } C_{313}) \text{ or } \cdots \\ \text{or } (C_{312} \text{ and } C_{313}) \text{ or } (C_{312} \text{ and } C_{314}) \text{ or } \cdots \tag{50}$$

The probabilities of such combinations are completely determined once we give an initial joint probability distribution for all statements (48) and their negations, which is equivalent to the joint distribution of

$$\mathrm{p}[(c_{mij}) \mid B] \equiv \mathrm{p}(c_{111}, c_{112}, c_{113}, \dots \mid B) \tag{51}$$

conditional on the information B that we have before doing the experiments and the observations.

Now, we'll want the probability of a specific answer A to our question, conditional on the data D we observed: $\mathrm{p}(A \mid D, B)$. By the rule of conditional probability we obtain it as

$$\mathrm{p}(A \mid D \text{ and } B) = \frac{\mathrm{p}(A \text{ and } D \mid B)}{\mathrm{p}(D \mid B)}. \tag{52}$$

The probabilities in the numerator and denominator can be obtained from our main joint probability (51) using the basic rules of the probability-calculus.

So our strategy is as follows:

1. Quantitatively determine the initial joint probability $\mathrm{p}[(c_{mij}) \mid B]$.
2. Resolve the answers to question Q2 in terms of the basic statements C_{mij}.
3. Resolve the experimental observations in terms of the basic statements C_{mij}.
4. Calculate the probabilities in (52).

We now face these steps in turn. We shall discuss several possible alternative assumptions and conceptual or mathematical approximations that we can make in order to proceed.

5 Initial probability distribution

In order to formulate the initial probability distribution $\mathrm{p}[(c_{mij}) \mid B]$ we explore the symmetries of our problem.

Let's analyse two approximate symmetries first. Consider a specific mouse m.

For each neuron i in the input region, we can approximately judge as equally likely that it will connect to any of the neurons j in the target region. This means that our initial probability distribution should remain

the same if we swap some values of the index j, with m and i fixed. For example, the probability that neuron #3 in the input region projects to neuron #2 in the target region and does not project to neuron #5 should be the same if we exchange 2 and 5 in that statement. And similarly for combinations of such statements:

$$\mathrm{p}(c_{m\,3\,2}=1, c_{m\,3\,5}=0 \mid B) = \mathrm{p}(c_{m\,3\,2}=0, c_{m\,3\,5}=1 \mid B). \tag{53}$$

This symmetry is an approximation: if the indices i and j bear some information about the neurons' locations, our probability should reflect the possibility that nearby neurons may connect more easily, for example.

We have an analogous approximate symmetry under exchanges of the input-neuron index i for fixed target neuron j.

These two approximate symmetries together are important because they greatly restrict the possible values of our initial probability distribution $\mathrm{p}[(c_{mij}) \mid B]$. The result they lead to[12] is mathematically difficult for me, however. To further simplify the problem I make an even rougher approximation: that the symmetries above hold not just for i and j separately, but for any pairs of them. This assumption is even more unrealistic than the previous two, because it gives the same probability to the situation where each input neuron is connected with only one, distinct, target neuron, as to the situation where only one input neuron is connected to all target neurons, and all other input neurons are unconnected.

With this approximation the joint probability distribution for a given mouse m assumes the following form, by de Finetti's representation theorem:[13]. Take a set of k_m input neurons and l_m target neurons, and consider all possible connections c_{mij} between them, with $i = 1, \dots, k_m$ and $j = 1, \dots, l_m$. Then

$$\begin{aligned}\mathrm{p}[(c_{mij}) \mid m, B] &= \int_0^1 \mathrm{d}\nu_m \; \mathrm{p}(\nu_m \mid B) \; \prod_{ij} \left[{\nu_m}^{c_{mij}} \, (1-\nu_m)^{1-c_{mij}} \right] \\ &\equiv \int_0^1 \mathrm{d}\nu_m \; \mathrm{p}(\nu_m \mid B) \; {\nu_m}^{k_m l_m f_m} \, (1-\nu_m)^{k_m l_m (1-f_m)}.\end{aligned} \tag{54}$$

Here $k_m l_m f_m = \sum_{ij} c_{mij}$ is the total number of connections between the k_m input neurons and the l_m target neurons, and $0 \leqslant \nu_m \leqslant 1$ is the

[12] Hoover 1979; Aldous 1981; Diaconis et al. 1981.

[13] De Finetti 1930; Hewitt et al. 1955; Heath et al. 1976; Diaconis 1977; Diaconis et al. 1980; Dawid 2013.

population-averaged number of connections from the *whole* input region to the *whole* target region. The probability distribution $\mathrm{p}(\nu_m \mid I)$ expresses our beliefs about the total number of connections from input to target region in mouse m. The formula above is valid only if k_m and l_m are small (say, one tenth at most) compared to their region sizes.

The probability distribution for the connections present in different mice can be obtained by combining the expressions above for each m. To do this we need the initial joint probability distribution

$$\mathrm{p}[(\nu_m) \mid B] \equiv \mathrm{p}(\nu_1, \nu_2, \dots \mid B) \tag{55}$$

for the total number of connections in several mice.

Our question has another symmetry, more realistic than the previous two. We don't believe any of the mice to be examined to be special with respect to the others, nor we do believe that the particular order in which they are observed has any relevance. This means that our initial probability distribution above should be invariant with respect to permutations of the index m. Then we obtain the following expression, again by de Finetti's theorem:

$$\mathrm{p}[(\nu_m) \mid B] = \int \mathrm{d}\theta \; \mathrm{p}(\theta \mid B) \prod_m q(\nu_m \mid \theta), \tag{56}$$

where q is a distribution that depends on the parameter θ, and $\mathrm{p}(\theta \mid B)$ is our initial probability distribution for this parameter.

The parameter θ in principle represents every possible distribution – or histogram – of the total number of connections across the mice. For example, a particular θ can say that 7% of mice had 10^6 connections, 11% had 5×10^7 connections, and so on. The above formula sums over all such possible distributions. There is an infinite of such distributions, so the integral is over an infinite-dimensional space. When we calculate this integral with this full generality, we're using a so-called *non-parametric* method. But for mathematical simplicity we can also decide to restrict our attention to histograms of specific shape, such as a normal for example. In this case the integral is restricted to a finite-dimensional space and we're using a *parametric* method.

Combining the initial probability distributions (54) and (56) we finally obtain our initial probability distribution for a joint set of statements

(C_{mij}):

$$\mathrm{p}[(c_{mij}) \mid B] = \iint \mathrm{d}\theta \; \mathrm{d}\boldsymbol{\nu} \; \mathrm{p}(\theta \mid B) \; \prod_{mij} \Big[q(\nu_m \mid \theta) \; {\nu_m}^{c_{mij}} \; (1-\nu_m)^{1-c_{mij}} \Big], \tag{57}$$

where $\boldsymbol{\nu} \coloneqq (\nu_m)$.

6 Statements of the questions and of the data

Bibliography

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

Aldous, D. J. (1981): *Representations for partially exchangeable arrays of random variables*. J. Multivariate Anal. **11**[4], 581–598.

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.

— (1981): *On the statistics of vision: the Julesz conjecture*. J. Math. Psychol. **24**[2], 112–138. http://statweb.stanford.edu/~cgates/PERSI/year.html.

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.

Hoover, D. N. (1979): *Relations on probability spaces and arrays of random variables*. Tech. rep. (Institute for Advanced Study, Princeton). http://www.stat.berkeley.edu/~aldous/Research/hoover.pdf.

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/logic03john.

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