

# Inferences from a network to a subnetwork and vice versa under an assumption of symmetry

P.G.L. Porta Mana\* E. Torre V. Rostami

*Institute of Neuroscience and Medicine (INM-6),  
Institute for Advanced Simulation (IAS-6), JARA BRAIN Institute I,  
Forschungszentrum Jülich, Germany*

\* (p.mana@fz-juelich.de)

10 December 2015

This note summarizes some mathematical relations between the probability distributions for the states of a network of binary units and a subnetwork thereof, under an assumption of symmetry. These relations are standard results of probability theory, but seem to be rarely used in neuroscience. Some of their consequences for inferences between network and subnetwork, especially in connection with the maximum-entropy principle, are briefly discussed. The meanings and applicability of the assumption of symmetry are also discussed.

Keywords: neuronal networks, statistical models, maximum-entropy

PACS: 87.19.L-,87.19.l0,05.90.+m MSC: 03B48,97K50

Remark: In this Note, whenever we use the term “probability”, we are not speaking about a frequency, or about some sort of (unmeasurable) physical property, but about a degree of reasonable belief, or plausibility, whose manipulation follows a well-defined and well-established logical calculus, “Bayesian theory” [1–27].

**probable**, *a.* and *n.*

**3. a.** Having an appearance of truth; that may in view of present evidence be reasonably expected to happen, or to prove true; likely.

Oxford English Dictionary [28]

The correct probability is always that relative to the knowledge of the person who makes the probability-judgment.

C. D. Broad [29, ch. II, p. 150]

When people say that the proposition “it is probable that  $p$  will occur” says something about the event  $p$ , they forget that the probability remains even when the event  $p$  does *not* occur.

L. Wittgenstein [4, p. 227]

## 1. Uncertainties about networks and subnetworks

If we are uncertain about the state of a network of neurons, what is our uncertainty about the state of a subnetwork? And vice versa, if we are uncertain about the state of a subnetwork, what is our uncertainty about the state of the whole network?

If our uncertainties are expressed by two probability distributions for the states of network and subnetwork, and these distributions satisfy a particular symmetry property, then they are related by precise and relatively simple mathematical formulae. These formulae are of essential importance when we want to make inferences about the whole network given data about the subnetwork and vice versa.

While well-known in survey sampling and in the pedagogic problem of “drawing from an urn without replacement”, such formulae are somewhat hard to find explicitly written in the neuroscientific literature. We have therefore decided to summarize them in this Note, as a reference, for the case of neurons modelled as “units” with binary  $\{0, 1\} \equiv \{\text{inactive, active}\}$  states. The formulae also apply to any network or population of any kind of units modelled in the same way: e.g., the values of the physical connectivities of a neuronal network.

The formulae and the assumptions underlying them are given in the next two sections, followed by examples of their use in both inferential directions: network to subnetwork and reverse. We also discuss the meanings of the symmetry property, and conclude with a brief summary. Proofs of the formulae are sketched in an appendix.

Our notation follows ISO [30], ANSI/IEEE [31], and NIST [32] standards but for the use of the comma “,” to denote logical conjunction (“and”, “ $\wedge$ ”), for the sake of typographic poise. Probability-calculus notation follows Jaynes’s book [22].

## 2. Formulae connecting network and subnetwork probabilities

### 2.1. Setup

Consider a network of  $N$  binary neurons with states  $(X_1, \dots, X_N)$  having fixed but unknown binary values  $(R_1, \dots, R_N)$ , with  $R_i$  in  $\{0, 1\}$ , vectorially written  $\mathbf{X} = \mathbf{R}$ . For example,  $\mathbf{X}$  can represent the state of a network at a particular time. We call the neurons “units” to lend some generality to

our discussion. We shall make statements about the whole network of  $N$  units and about a subnetwork of  $n$  units; the word “network” will always refer to the *whole* network. The subnetwork states and their values are denoted by lowercase letters:  $(x_1, \dots, x_n) \equiv \mathbf{x}$  and  $(r_1, \dots, r_n) \equiv \mathbf{r}$ ; but note that  $x_i \equiv X_{j_i}$  and  $r_i \equiv R_{j_i}$  for some distinct  $j_1, \dots, j_n$ . We shall also make statements about the network-averaged state, or *network average*:

$$\bar{X} := (X_1 + \dots + X_N)/N, \quad (1)$$

and the subnetwork-averaged state, or *subnetwork average*:

$$\hat{x} := (x_1 + \dots + x_n)/n. \quad (2)$$

The quantities  $N\bar{X}$  and  $n\hat{x}$  represent the total number of active units in the network and the subnetwork. Quantities like  $\bar{R}$  and  $\hat{r}$  are defined analogously. The averaging operators  $\widehat{\cdot}$  and  $\bar{\cdot}$  are also extended to averages of  $\binom{n}{m}$  or  $\binom{N}{m}$  products of  $m$  states; e.g.,

$$\bar{X X} := \binom{N}{2}^{-1} (X_1 X_2 + X_1 X_3 + \dots + X_{N-1} X_N), \quad (3)$$

$$\widehat{x x x} := \binom{n}{3}^{-1} (x_1 x_2 x_3 + x_1 x_2 x_4 + \dots + x_{n-2} x_{n-1} x_n), \quad (4)$$

and so on.

## 2.2. Assumptions

Our uncertainty about the network state is represented by the joint probability distribution of the individual states, from which we can derive all other probabilities of interest. We denote it by

$$P(X_1 = R_1, X_2 = R_2, \dots, X_N = R_N | H) \quad \text{or} \quad P(\mathbf{X} = \mathbf{R} | H). \quad (5)$$

Such probability is conditional on our state of knowledge, i.e. the evidence and assumptions backing our probability assignments, denoted by the proposition  $H$ .

In the present discussion,  $H$  is a state of knowledge that leads to two specific properties in our probability assignments:

*H1. Permutation symmetry*, expressed as the invariance of the joint distribution (5) under arbitrary permutations of the units’s labels:

$$P(X_1 = R_1, X_2 = R_2, \dots, X_N = R_N | H) = P(X_1 = R_{\pi(1)}, X_2 = R_{\pi(2)}, \dots, X_N = R_{\pi(N)} | H) \quad \text{for any permutation } \pi. \quad (6)$$

This property can reflect two very different states of knowledge: physical homogeneity of the network, or symmetry in our ignorance about the network. This property is called *finite exchangeability* in the Bayesian literature and its basis, consequences, and alternatives to it are discussed in § 4.

*H2.* The network average  $\bar{X}$  has a particular distribution  $Q$ :

$$P(\bar{X} = A | H) = Q(A), \quad A \in \{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}. \quad (7)$$

For the moment we are not concerned about the specific form of  $Q$  and about how it was assigned: it could, e.g., arise from maximum-entropy arguments [e.g.: 33–42] used with data on the network.

### 2.3. Formulae

The state of knowledge  $H$  has the following six (not independent) main consequences for our probability assignments:

*I.* Probability for the network state:

$$P(X = R | H) = \binom{N}{N\bar{R}}^{-1} Q(\bar{R}). \quad (8)$$

*II.* Probability for the state  $x$  of any subnetwork of  $n$  units:

$$P(x = r | H) = \sum_{NA=0}^N \binom{N-n}{NA-n\widehat{r}} \binom{N}{NA}^{-1} Q(A). \quad (9)$$

Note that the only summands contributing to this sum are those for which  $n\widehat{r} \leq NA \leq N$ ; the others are zero because by definition  $\binom{M}{y} = 0$  if  $y < 0$ . This remark applies to all the sums of this kind in the rest of this Note.

*III.* Probability for the subnetwork state conditional on a network state:

$$P(x = r | X = R, H) = \binom{N-n}{N\bar{R}-n\widehat{r}}. \quad (10)$$

#### IV. Probability for the subnetwork average $\widehat{x}$ :

$$P(\widehat{x} = a | H) = \binom{n}{na} \sum_{NA=0}^N \binom{N-n}{NA-na} \binom{N}{NA}^{-1} Q(A),$$

$$a \in \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\right\}. \quad (11)$$

V. Probability for the subnetwork average conditional on the network average:

$$P(\widehat{x} = a | \bar{X} = A, H) = \binom{n}{na} \binom{N-n}{NA-na} \binom{N}{NA}^{-1}. \quad (12)$$

VI. The product of the states of any  $m$  distinct units from a given subnetwork,

$$x_{i_1} x_{i_2} \cdots x_{i_m}, \quad 1 \leq i_1 < i_2 < \cdots < i_m \leq n$$

has an expectation equal to that of the subnetwork average of such products, is independent of the subnetwork size  $n$ :

$$E(x_{i_1} \cdots x_{i_m} | H) = E(\underbrace{\widehat{x} \cdots \widehat{x}}_{m \text{ factors}} | H) = E(\underbrace{\bar{X} \cdots \bar{X}}_{m \text{ factors}} | H), \quad (13a)$$

and has an explicit expression in terms of  $Q$ :

$$E(x_{i_1} \cdots x_{i_m} | H) = \binom{N}{m}^{-1} \sum_{NA=0}^N \binom{NA}{m} Q(A)$$

$$\equiv \sum_{NA=0}^N \binom{N-m}{NA-m} \binom{N}{NA}^{-1} Q(A). \quad (13b)$$

A useful relation connects the expectation of a product (13) and the  $m$ th factorial moment [43] of the probability distributions for the averages. The  $m$ th factorial moment of the subnetwork average  $\widehat{x}$  is defined by

$$E[\underbrace{n\widehat{x}(n\widehat{x}-1)\cdots(n\widehat{x}-(m-1))}_{m \text{ factors}} | H] \equiv E\left[\frac{(n\widehat{x})!}{(n\widehat{x}-m)!} | H\right], \quad (14)$$

an analogous definition holding for  $\bar{X}$ . We have that

$$E(x_{i_1} \cdots x_{i_m} | H) = \frac{(n-m)!}{n!} E\left[\frac{(n\widehat{x})!}{(n\widehat{x}-m)!} | H\right] = \frac{(N-m)!}{N!} E\left[\frac{(N\bar{X})!}{(N\bar{X}-m)!} | H\right]. \quad (15)$$

As a consequence of the above relation, the first three moments of the probability distributions  $P(\widehat{x} = a | H)$  and  $P(\overline{X} = A | H)$ , are related by

$$E(\widehat{x} | H) = E(\overline{X} | H), \quad (16a)$$

$$E(\widehat{x}^2 | H) = E(\overline{X} | H) \frac{N-n}{(N-1)n} + E(\overline{X}^2 | H) \frac{N(n-1)}{(N-1)n}, \quad (16b)$$

$$E(\widehat{x}^3 | H) = E(\overline{X} | H) \frac{(N-n)(N-2n)}{(N-1)(N-2)n^2} + E(\overline{X}^2 | H) \frac{3N(N-n)(n-1)}{(N-1)(N-2)n^2} + E(\overline{X}^3 | H) \frac{N^2(n-1)(n-2)}{(N-1)(N-2)n^2}. \quad (16c)$$

Relations for higher moments can be obtained recursively from eq. (15). In general, this means that the two sets of first  $m$  moments are related by a homogeneous linear transformation,

$$E(\widehat{x}^m | H) = \sum_{l=1}^m M_{ml}(n, N) E(\overline{X}^l | H), \quad (17)$$

with a universal, lower-triangular transformation matrix  $M_{ml}(n, N)$  that depends only on  $n, N$ , and the condition of symmetry (6).

As intuition suggests, we have

$$E(\widehat{x}^m | H) \xrightarrow{n \rightarrow N} E(\overline{X}^m | H), \quad E(\widehat{x}^m | H) \xrightarrow{n \rightarrow 1} E(\overline{X} | H), \quad (18)$$

the latter because  $x_i^m = x_i$ , since states are  $\{0, 1\}$ -valued.

The core of the six mathematical relations above are eqs (9) and (11). The latter expresses the probability for the subnetwork average as a mixture of hypergeometric distributions [22, ch. 3; 44, § 4.8.3; 45, § II.6], with parameters  $N, N\overline{X}, n$ , weighted by the probabilities  $P(\overline{X} = A | H)$  [cf. 46, § 4, esp. eq. (22)]. The connection between this mixture representation and the condition of symmetry (6) is well-known in the Bayesian literature [46–51].

Proofs of the above formulae are sketched in appendix A.

## 2.4. Asymptotic approximations and generalizations

Recall the definition of the Shannon and Burg [52; 53] entropies for a binary probability distribution  $(p, 1 - p)$  with  $p$  in  $[0, 1]$

$$H(p) := -p \ln p - (1 - p) \ln(1 - p), \quad H_B(p) := \ln p + \ln(1 - p). \quad (19)$$

The binomial coefficient has an asymptotic form that involves the two entropies above [54–60]:

$$\ln \binom{M}{Mp} = M H(p) - \frac{1}{2} \ln(2\pi M) - \frac{1}{2} H_B(p) + O\left(\frac{1}{\sqrt{M}}\right), \quad M \text{ large}, \quad (20)$$

and this expression can be used to obtain asymptotic forms of the mathematical formulae (8)–(15) for large  $N$ , depending on the magnitudes of  $n$  and  $n/N$ . For example, if  $N$  and  $n$  are large and their ratio  $k := n/N$  finite, the probability distributions for the averages can be approximated (provided some regularity conditions) by continuum densities,

$$n P(\hat{x} = a | H) \approx f(a), \quad N P(\bar{X} = A | H) = N Q(A) \approx F(A), \quad (21)$$

$$f, F: [0, 1] \rightarrow [0, +\infty[,$$

and sums by integrals, and eq. (9) takes on the approximate form

$$f(a) = \int_0^1 \exp\left\{N \left[ k H(a) - H(A) + (1 - k) H\left(\frac{A - ka}{1 - k}\right) \right]\right\} F(A) dA, \quad (22)$$

which shows an interplay of three entropies that depends on the ratio  $n/N$ . In some cases this allows us to approximate the integral by the value at a mode determined by a generalized entropy principle. (In fact, the standard maximum-entropy procedure can be derived from the probability calculus via a similar approximation [61].)

The relations above can also be generalized to  $K$ -valued states  $X_i$  in  $\{0, \dots, K - 1\}$ , leading to the appearance of the generalized hypergeometric distribution [22, ch. 3; 44, § 4.8.3; 45, § II.6], or to real-valued states  $X_i$  in  $\mathbf{R}$ .

We do not pursue any of these approximations or generalizations in this Note.

### 3. Examples of inferential use of the formulae

#### 3.1. From network to subnetwork

Let us illustrate with an example how the probability distribution for the subnetwork average  $\hat{x}$ , determined by eq. (11), changes with the subnetwork size  $n$ . Choose a network-average distribution  $P(\bar{X} = A | H)$  belonging to the exponential family [62, § 4.5.3; see also 63]:

$$P(\bar{X} = A | H) = Q(A) \propto \binom{N}{NA} \exp[\lambda_2 NA (NA - 1)/2 + \lambda_1 NA]. \quad (23)$$

This is the form obtained from the principle of maximum relative entropy [e.g.: 33–42] with first and second moments as constraints and the reference distribution  $Q_0$  defined by  $Q_0(A) = 2^{-N} \binom{N}{NA}$ , corresponding to a uniform probability distribution for the network state  $X$ .

The probability distribution of eq. (23) is plotted in fig. 1, together with the resulting subnetwork-average distributions  $P(\hat{x} = a | H)$ , for the case in which  $N = 1000$  units,  $\lambda_1 = -2.55$ ,  $\lambda_2 = 0.005$ , and  $n = 10, 50, 100, 250$ . The distributions become broader as  $n$  decreases, and the minimum of the original distribution disappears; at the same time the finite-difference

$$\frac{P(\hat{x} = a + 1/n | H) - P(\hat{x} = a | H)}{1/n}$$

presents a sharp jump at this minimum when  $n \approx 100$ .

To the eye familiar with maximum-entropy distributions, the subnetwork-average distributions of fig. 1 do not look like maximum-entropy ones with second-moment constraints. In fact, they are not and *cannot* be:

$$P(\hat{x} = a | H) \neq \kappa \binom{n}{na} \exp[\kappa_2 na (na - 1)/2 + \kappa_1 na] \quad (24)$$

for any  $\kappa, \kappa_1, \kappa_2$ , unless  $n = 2$ . This impossibility holds more generally for any number of constraints  $m$  and subnetwork size  $n$  such that  $m < n$ . The reason is simple: suppose we have assigned a maximum-entropy distribution with  $m$  moment constraints as the distribution for the network average. If we want the same kind of distribution for a subnetwork of size  $n$ , we are free to play with  $m + 1$  parameters (normalization included), but we must also satisfy the  $n + 1$  equations corresponding to the marginalization (11). This is generally impossible

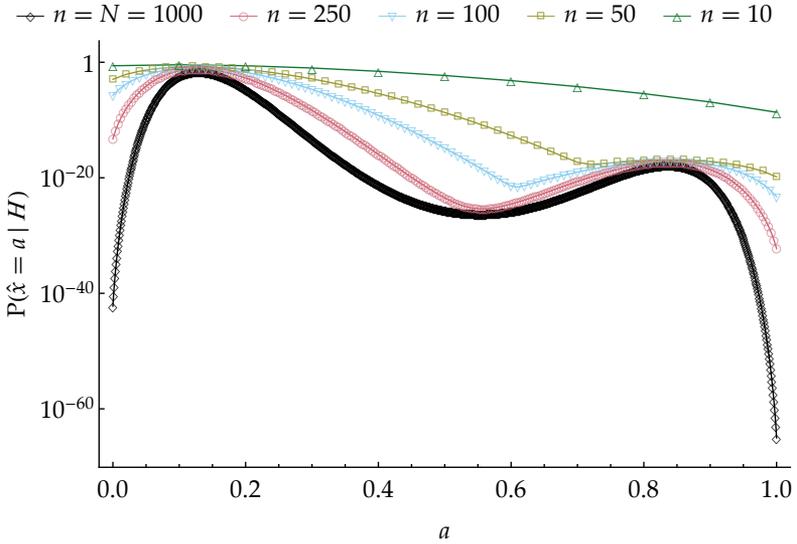


Figure 1 Probability distributions  $P(\hat{x} = a | H)$  for different subnetwork sizes  $n$ , obtained from a network probability distribution  $P(\bar{X} = A | H)$  having the maximum-entropy form (23).

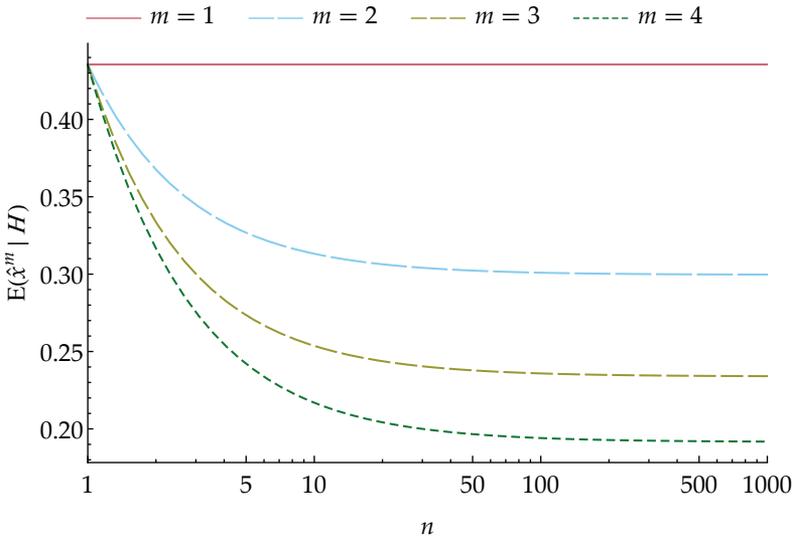


Figure 2 Moments of the probability distributions  $P(\hat{x} = a | H)$  as functions of the subnetwork size  $n$ .

unless  $m \geq n$ . (Impossibilities of a similar kind appear in statistical mechanics, see e.g. ref. [64].)

This fact can be significant for recent works [e.g., 65–73] in which a maximum-entropy probability distribution with second- or third-moment constraints is assigned to relatively small subnetworks ( $n < 200$ ) of neurons. If we assume that such subnetwork is part of a larger network, and assume the condition of symmetry (6), then the larger network *cannot* be assigned a maximum-entropy distribution with the same number of constraints. Vice versa, if we assign such a maximum-entropy distribution to the larger network, then none of its subnetwork of enough large size  $n$  can be assigned a similar maximum-entropy distribution. See ref. [74] for a broader discussion of this fact and of its consequences.

The dependence of the first four moments  $E(\hat{x}^m | H)$  as a function of size  $n$  is shown in fig. 2. The moments become practically constant when  $n \approx 100$  or larger. The expectations of  $m$ -tuple products of states  $E(x_{i_1} \cdots x_{i_m} | H)$ , proportional to the factorial moments, are not shown as they do not depend on  $n$ .

### 3.2. From subnetwork to network

We have seen that, given the condition of symmetry (6), the probability  $P(\bar{X} = A | H)$  for the network average determines that of each subnetwork average,  $P(\hat{x} | H)$ , by the marginalization eq. (11). The reverse is trivially not true, since eq. (11), as a linear mapping from  $\mathbf{R}^{N+1}$  to  $\mathbf{R}^{n+1}$ , with  $N$  larger than  $n$ , is onto but not into. Assigning a probability distribution  $P(\hat{x} = a | H)$  to a subnetwork average  $\hat{x}$  does not determine a network distribution  $P(\bar{X} = A | H)$ : it only restricts the set of possible ones; this set can in principle be determined via linear-programming methods [23; 25; 75–77].

Analogous situations appear in the truth-valued logical calculus: if the composite proposition  $A \Rightarrow B$  is assigned the truth-value “true”, then assigning  $A$  the value “true” also determines the value of  $B$ , whereas assigning  $B$  the value “true” leaves the value of  $A$  undetermined.

The same linear-programming methods show that any inference from subnetwork properties to network ones must necessarily start from some assumptions  $I$  that assign a probability distribution  $P(X = \mathbf{R} | I)$  for the network states. The approaches to this task and reformulations of it have become uncountable: they include exchangeable models, parametric and non-parametric models, hierarchical models, general linear models,

models via sufficiency, maximum-entropy models, and whatnot [e.g.: 5; 22; 62; 78–86]. We now show two examples, based on a maximum-entropy approach, that to our knowledge have not yet been explored in the neuroscientific literature. For a concrete application see [87].

**First example: moment constraints for the network.** Consider a state of knowledge  $H'$  leading to the following properties:

$H'1.$  the expectations of the single and pair averages  $\widehat{x}$  and  $\widehat{xx}$  of a particular subnetwork have given values

$$E(\widehat{x} | H') = c_1, \quad E(\widehat{x_i x_j} | H') = c_2; \quad (25)$$

$H'2.$  the network probability distribution  $P(X = R | H')$  has maximum relative entropy with respect to the uniform one, given the constraints above.

Then the probability distribution for the network conditional on  $H'$  is completely determined: it satisfies the symmetry property (6) and is defined by

$$p(X = R | H') = K \exp[\Lambda_2 N \bar{R} (N \bar{R} - 1)/2 + \Lambda_1 N \bar{R}]$$

with  $K, \Lambda_m$ , such that the distribution is normalized and

$$K \sum_{NA=0}^N \binom{N-m}{NA-m} \exp[\Lambda_2 NA (NA-1)/2 + \Lambda_1 NA] = c_m, \quad m = 1, 2. \quad (26)$$

We omit the full proof of this statement: it is a standard application of the maximum-entropy procedure [e.g.: 33–36; 38–42], combined with the equality (13) of subnetwork and network expectations, e.g.

$$c_2 = E(\widehat{xx} | H') = \binom{N}{2}^{-1} \sum_{NA=0}^N \binom{NA}{2} P(\bar{X} = A | H'), \quad (27)$$

and with relations (8), (10). This example is easily generalized to any number  $m$  of constraints such that  $m \leq n$ .

Note again that, as remarked in § 3.1, the subnetwork from which the averages in the expectations (25) are calculated has a probability distribution  $P(\widehat{x} = a | H)$  determined by the marginalization (11) and does *not* have a maximum-entropy form with the same number of constraints.

**Second example: subnetwork-distribution constraint.** Consider another state of knowledge  $H''$  leading to the following properties:

$H''$ 1. the average  $\widehat{x}$  of a particular subnetwork has a probability distribution  $q$ :

$$P(\widehat{x} = a | H'') = q(a); \quad (28)$$

$H''$ 2. the probability distribution for the network,  $P(X = R | H'')$ , has maximum relative entropy with respect to the uniform one, given the constraint above.

Then the probability distribution for the network given  $H''$  is completely determined and satisfies the symmetry property (6):

$$P(X = R | H'') = \exp \left[ \sum_{na=0}^n \Lambda_a \binom{n}{na} \binom{N-n}{N\bar{R}-na} \right]$$

with  $\Lambda_a$  such that

$$\sum_{NA=0}^N \binom{n}{na} \binom{N-n}{NA-na} \exp \left[ \sum_{na=0}^n \Lambda_a \binom{n}{na} \binom{N-n}{NA-na} \right] = q(a) \quad (29)$$

(the normalization constraint being unnecessary since  $q$  is normalized). This result is just another application of the maximum-entropy procedure with  $n + 1$  (linear) constraints given by eq. (9), where the left-hand side is now given and equal to  $q(a)$ .

This example is equivalent to the generalization of the previous one with  $n$  moment constraints, since knowledge of  $P(\widehat{x} = a | H'')$  is equivalent to knowledge its first  $n$  moments.

The above examples do not mention how the values of the expectation constraints or of the subnetwork-average probability distribution can have been assigned. They cannot be assigned by a measurement, of course, since probabilities and expectations are not physical quantities and cannot be physically measured – they represent guesses of an observer and depend on the observer’s state of knowledge and assumptions. Rather, such values usually come from measurements made on “copies” – in a very general sense – of the states of the network; e.g., when we have a time sequence of them and measure the frequencies of their occurrence. Such situations can again be fully analysed with the probability calculus, and one can show [61] that the maximum-entropy formulae in the examples above are just limit forms of such an analysis,

employing measured physical data  $\Delta$  like, e.g., frequencies, and repeated applications of Bayes's theorem,

$$P(X = R | \Delta, I) \propto P(\Delta | X = R, I) P(X = R | I), \quad (30)$$

which updates the initial probability assignments on such data. But the discussion of this is again outside the scope of this Note.

#### 4. On the symmetry property

The symmetry property (6) is called *finite exchangeability* in the Bayesian literature and, as was mentioned in § 2.3, its relation to the hypergeometric distribution in expression (11) for the subnetwork average is well-known [46–51].

This property can reflect two very different states of knowledge: either (a) knowledge that the network is somehow *physically homogeneous*, or (b) *complete ignorance* about the network's homogeneity or inhomogeneity. In the second case we are saying that the indices or labels “*i*” of the units are *uninformative* – because, for example, we have no idea of how the units were labelled, hence we cannot presuppose any relation among the units, nor can we presuppose any structural or topological properties of the network they constitute. In the first case we are saying that the labels are *irrelevant*, even though they might be informative. The reason could be that, even if there is a connection between labels and, say, spatial locations of the units, each unit is nevertheless physically, homogeneously, identically connected to all the others; so we assume that spatial location does not play a relevant role.

An important consequence of the symmetry property is that no amount of new evidence about an *asymmetry* in the labelling of some units can lead to asymmetric predictions about the remaining ones. For example, if we have new data  $\Delta$  saying that the first  $n$  units are in state 1 and the last  $n$  in state 0,

$$\Delta := (X_1 = X_2 = \dots = X_n = 1, \quad X_{N-(n-1)} = \dots = X_{N-1} = X_N = 0), \quad (31)$$

with  $n$  large, our updated probability distribution will still assign (as can be shown using eqs (6) and (8)) the same probability for their respective

neighbouring units with labels  $n + 1$  and  $N - n$  to be in state 1 or 0:

$$\begin{aligned} P(X_{n+1} = 1 | \Delta, H) &= P(X_{N-n} = 1 | \Delta, H), \\ P(X_{n+1} = 0 | \Delta, H) &= P(X_{N-n} = 0 | \Delta, H). \end{aligned} \quad (32)$$

This may seem unreasonable – we would now say that the unit  $n + 1$  is more likely to be in state 1, as the first  $n$  are, than in 0; and that the unit  $N - n$  is more likely in state 0, as the last  $n$  are, than in 1; i.e.

$$\begin{aligned} P(X_{n+1} = 1 | \Delta, H) &> P(X_{n+1} = 0 | \Delta, H), \\ P(X_{N-n} = 1 | \Delta, H) &< P(X_{N-n} = 0 | \Delta, H), \end{aligned} \quad (33)$$

which is equivalent to  $P(X_{n+1} = 1 | \Delta, H) > \frac{1}{2}$ ,  $P(X_{N-n} = 1 | \Delta, H) < \frac{1}{2}$ ; but the equalities (32) make this impossible.

The two very different motivations behind the symmetry property – lack of information and lack of relevance – can of course be handled differently by the probability calculus, in such a way that the appearance of asymmetry in new data leads to asymmetry in updated predictions. But this requires a more complex set of assumptions than those embodied in  $H$ ; it requires, in particular, some sort of probability distribution for the degree of physical symmetry or homogeneity of the network, appropriately quantified.

The moral is that we should use the symmetric assumption  $H$  only if we can safely exclude the presence of inhomogeneity or are uninterested in detecting its presence. Otherwise, we must resort to more appropriate (and complex) assumptions. If we repeatedly observe new values that happen to have a very low probability according to the updated distribution  $P(X = R | \Delta_1, \Delta_2, \dots, H)$ , this could be an indication that the symmetry property is unreasonable. Any strong departure of higher powers of measured averages from their expected behaviour given in eq. (16) and illustrated in fig. 2 can also be an indication that the symmetry property may have to be abandoned; hence the usefulness of eq. (16).

## 5. Summary and remarks

The main point of this Note was to explicitly collect and illustrate the mathematical formulae I–VI, § 2.3, between the probabilities assigned to the state of a network of neurons (or similar entities), and those assigned

to the state of a subnetwork thereof. The formulae hold in the simple case of binary states and under an assumption of symmetry. Such relations can be found in several classical texts on probability and statistics; but we deemed it useful to restate them in a neuroscientific context, given their fundamental importance in relating a whole to its parts.

We have indeed seen that these formulae lead to straightforward but in some cases unexpected consequences: e.g., if a network is assigned a maximum-entropy probability distribution, then its subnetworks *cannot* be assigned a maximum-entropy distribution of the same form, and vice versa. The formulae also readily suggest new ways of making inferences or of formulating starting assumptions. Discussion of these points is left to forthcoming works [74; 87].

We also hope to have given the readers a feeling of the agility and simplicity with which the probability calculus (“Bayesian theory”) leads us from assumptions to consequences (albeit sometimes with non-simple mathematics), and from consequences to the assessment of how suitable our initial assumptions are, as shown with the assumption of symmetry.

## A. Sketched proofs

Variants of the following derivations and combinatorial considerations can be found e.g. in [88, chs I–IV; 45, ch. II; 22, ch. 3]; see also [89].

To derive the joint probability distribution (8) from that for the network average (7), consider that if the network total is  $N\bar{X}$ , then  $N\bar{X}$  out of  $N$  units are active, and there are  $\binom{N}{N\bar{X}}$  possible states for which this can be true; therefore

$$P(X = R | H) = \binom{N}{N\bar{R}}^{-1} Q(\bar{R}). \quad (8)_r$$

An analogous reasoning for  $n$  and  $\hat{x}$  leads to an analogous equality,

$$P(x = r | H) = \binom{n}{n\hat{r}}^{-1} P(\hat{x} = \hat{r} | H), \quad (34)$$

for the subnetwork.

Let us next consider the probability  $P(\hat{x} = a | \bar{X} = A, H)$  for the subnetwork average  $\hat{x}$  conditional on the network average  $\bar{X}$ . There are

$\binom{N}{N\bar{X}}$  possible network states if the network average is  $\bar{X}$ , i.e. if  $N\bar{X}$  units are in state 1; the conditional probability of each is therefore  $1/\binom{N}{N\bar{X}}$ , owing to the symmetry assumption (6). Now consider the subnetwork of the first  $n$  units. The conditional probability of having  $n\hat{x}$  specific ones in state 1 is the sum of the probabilities of all states for which  $N\bar{X} - n\hat{x}$  of the remaining  $N - n$  units are in state 1; there are  $\binom{N-n}{N\bar{X}-n\hat{x}}$  such states, all equally probable. Finally, there are  $\binom{n}{n\hat{x}}$  possible ways, all equally probable, in which  $n\hat{x}$  of the first  $n$  units can be in state 1. In formulae,

$$\begin{aligned}
 P(\hat{x} = a | \bar{X} = A, H) &= \sum_r^{\hat{r}=a} \sum_R^{\bar{R}=A} P(x = r | X = R, H) P(X = R | \bar{X} = A, H), \\
 &= \sum_r^{\hat{r}=a} \sum_R^{\bar{R}=A} P(x = r | X = R, H) \binom{N}{NA}^{-1}, \\
 &= \sum_r^{\hat{r}=a} \binom{N-n}{NA-na} \binom{N}{NA}^{-1}, \\
 &= \binom{n}{na} \binom{N-n}{NA-na} \binom{N}{NA}^{-1},
 \end{aligned} \tag{35}$$

which is the conditional probability (12). Note that this is just a derivation of the hypergeometric distribution [22, ch. 3; 44, § 4.8.3; 45, § II.6], which describes the probability of, say, drawing a proportion of  $\hat{x}$  blue balls in  $n$  drawings without replacement from an urn with  $N$  balls, a fraction  $\bar{X}$  of which are blue.

The probability of a subnetwork average  $\hat{x}$  is then, by marginalization,

$$\begin{aligned}
 P(\hat{x} = a | H) &= \sum_{NA=0}^N P(\hat{x} = a | \bar{X} = A, H) P(\bar{X} = A | H), \\
 &= \sum_{NA=0}^N \binom{n}{na} \binom{N-n}{NA-na} \binom{N}{NA}^{-1} Q(A).
 \end{aligned} \tag{36}$$

which proves the subnetwork-average formula (11). This formula, combined with eqs (34) and (8), leads to the conditional probability (10).

The independence of the expectation of products of states from the subnetwork size is trivial by marginalization:

$$\begin{aligned} \sum_x x_1 \cdots x_m P(x = \mathbf{r} | H) &= \sum_{\mathbf{r}, \mathbf{R}} r_1 \cdots r_m P(x = \mathbf{r} | X = \mathbf{R}, H) P(X = \mathbf{R} | H), \\ &= \sum_{\mathbf{r}, \mathbf{R}} r_1 \cdots r_m \delta(R_1 - r_1) \cdots \delta(R_m - r_m) P(X = \mathbf{R} | H), \\ &= \sum_{\mathbf{R}} R_1 \cdots R_m P(X = \mathbf{R} | H). \end{aligned} \quad (37)$$

All such  $m$ -fold products have the same expectation by symmetry, therefore their subnetwork average will do, too, being an average of equal terms.

Now consider the sum of all distinct products of states of two units in the subnetwork:

$$x_1 x_2 + x_1 x_3 + \cdots + x_{n-1} x_n.$$

The terms in this sum are either 0 or 1. The non-vanishing ones are those with index pairs chosen from the  $n\hat{x}$  units of the subnetwork which are in state 1, and there are  $\binom{n\hat{x}}{2}$  such choices, so the sum above is equal to  $\binom{n\hat{x}}{2}$ . The sum has  $\binom{n}{2}$  terms, so their average is  $\binom{n\hat{x}}{2} / \binom{n}{2}$ . Generalizing the argument to products of  $m$  units, we have that

$$\widehat{x_{i_1} \cdots x_{i_m}} = \binom{n\hat{x}}{m} \binom{n}{m}^{-1}. \quad (38)$$

Then, using eq. (11),

$$\begin{aligned} E(\widehat{x_{i_1} \cdots x_{i_m}} | H) &= \frac{(n-m)!}{n!} E(m! \binom{n\hat{x}}{m} | H) \\ &\equiv \frac{(n-m)!}{n!} \sum_{na=0}^n m! \binom{na}{m} P(\hat{x} = a | H), \\ &= \frac{(n-m)!}{n!} \sum_{NA=0}^N Q(A) \left[ \sum_{na=0}^n m! \binom{na}{m} \binom{n}{na} \binom{N-n}{NA-na} \binom{N}{NA}^{-1} \right]. \end{aligned} \quad (39)$$

The expression in brackets is the  $m$ th factorial moment of the hypergeometric function, and is given by [43]

$$\sum_{na=0}^n m! \binom{na}{m} \binom{n}{na} \binom{N-n}{N\bar{X}-na} \binom{N}{N\bar{X}}^{-1} = m! \binom{n}{m} \binom{na}{m} \binom{N}{m}, \quad (40)$$

which combined with the previous equation yields the second line of eq. (13); its last equality comes from the identity

$$\binom{N}{M} \binom{M}{m} = \binom{N}{m} \binom{N-m}{M-m}, \quad (41)$$

easily derived by writing the binomial coefficients in terms of factorials. Finally, eqs (16), relating the moments of the distributions for subnetwork and network averages, is obtained from the definition of moments,

$$E(\widehat{x}^m | H) := \sum_{na=0}^n a^m P(\widehat{x} = a | H), \quad E(\overline{X}^m | H) := \sum_{NA=0}^N A^m P(\overline{X} = A | H), \quad (42)$$

replaced in the equalities for the factorial moments (15), by recursively solving in terms of the moments of the network distribution.

## Acknowledgements

We thank Moritz Helias for many inspiring discussions, encouragement, and support; and gratefully acknowledge partial support by the Helmholtz young investigator group VH-NG 1028, the Helmholtz Portfolio Supercomputing and Modeling for the Human Brain, and the EU Grant 604102 (Human Brain Project). PGLPM thanks Buster Keaton for filling life with awe, Miri & Mari for continuous encouragement and support, the Forschungszentrum librarians for their prompt support in acquiring references, and the developers and maintainers of L<sup>A</sup>T<sub>E</sub>X, Emacs, AUCT<sub>E</sub>X, MiK<sub>T</sub>E<sub>X</sub>, arXiv (except for its new “endorsement” system), Inkscape, Sci-Hub for making a free and unfiltered scientific exchange possible.

## Bibliography

- [1] P. S. Laplace (Marquis de): *Théorie analytique des probabilités*, 3rd ed. (Paris: Courcier, 1820). First publ. 1812; repr. in [90]; <http://gallica.bnf.fr/document?O=N077595>.
- [2] F. P. Ramsey: *Truth and probability*. In: [91] (1950), pp. 156–198. Repr. in [92], pp. 23–52; written 1926.
- [3] W. E. Johnson: *Probability: the deductive and inductive problems*. *Mind* **41**<sup>164</sup> (1932), 409–423. With some notes and an appendix by R. B. Braithwaite.

- [4] L. Wittgenstein: *Probability*. In: [93] (2004). Part I, App. 7.
- [5] H. Jeffreys: *Theory of Probability*, 3rd ed. (London: Oxford University Press, 2003). First publ. 1939.
- [6] B. O. Koopman: *The bases of probability*. Bull. Am. Math. Soc. **46** (1940), 763–774. Repr. in [92], pp. 159–172.
- [7] B. O. Koopman: *The axioms and algebra of intuitive probability*. Ann. Math. **41**<sup>2</sup> (1940), 269–292.
- [8] B. O. Koopman: *Intuitive probabilities and sequences*. Ann. Math. **42**<sup>1</sup> (1941), 169–187.
- [9] R. T. Cox: *Probability, frequency, and reasonable expectation*. Am. J. Phys. **14**<sup>1</sup> (1946), 1–13. [http : / / algomagic . org / ProbabilityFrequencyReasonableExpectation.pdf](http://algomagic.org/ProbabilityFrequencyReasonableExpectation.pdf).
- [10] G. Pólya: *Preliminary remarks on a logic of plausible inference*. Dialectica **3**<sup>1-2</sup> (1949), 28–35.
- [11] G. Pólya: *Mathematics and Plausible Reasoning: Vol. I: Induction and Analogy in Mathematics* (Princeton: Princeton University Press, 1954).
- [12] G. Pólya: *Mathematics and Plausible Reasoning: Vol. II: Patterns of Plausible Inference*, 2nd ed. (Princeton: Princeton University Press, 1968). First publ. 1954.
- [13] R. T. Cox: *The Algebra of Probable Inference* (Baltimore: The Johns Hopkins Press, 1961).
- [14] D. Scott, P. Krauss: *Assigning probabilities to logical formulas*. In: [94] (1966), pp. 219–264.
- [15] H. Gaifman: *Concerning measures in first order calculi*. Israel J. Math. **2** (1964), 1–18.
- [16] T. L. Fine: *Theories of Probability: An Examination of Foundations* (New York: Academic Press, 1973).
- [17] E. W. Adams: *The Logic of Conditionals: An Application of Probability to Deductive Logic* (Dordrecht: D. Reidel, 1975).
- [18] D. Lewis: *Probabilities of conditionals and conditional probabilities*. Phil. Rev. **85**<sup>3</sup> (1976), 297–315. See also [21].
- [19] H. Gaifman: *Subjective probability, natural predicates and Hempel's ravens*. Erkenntnis **14**<sup>2</sup> (1979), 105–147.
- [20] H. Gaifman, M. Snir: *Probabilities over rich languages, testing and randomness*. J. Symbolic Logic **47**<sup>3</sup> (1982), 495–548.
- [21] D. Lewis: *Probabilities of conditionals and conditional probabilities II*. Phil. Rev. **95**<sup>4</sup> (1986), 581–589. See also [18].

- [22] E. T. Jaynes: *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press, 2003). Ed. by G. Larry Bretthorst; <http://omega.albany.edu:8008/JaynesBook.html>, <http://omega.albany.edu:8008/JaynesBookPdf.html>, <http://www-biba.inrialpes.fr/Jaynes/prob.html>; first publ. 1994.
- [23] T. Hailperin: *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications* (London: Associated University Presses, 1996).
- [24] E. W. Adams: *A Primer of Probability Logic* (Stanford: CSLI, 1998).
- [25] T. Hailperin: *Logic with a Probability Semantics: Including Solutions to Some Philosophical Problems* (Plymouth, UK: Lehigh University Press, 2011).
- [26] A. Hájek: *What conditional probability could not be*. *Synthese* **137**<sup>3</sup> (2003), 273–323.
- [27] H. Gaifman: *Reasoning with bounded resources and assigning probabilities to arithmetical statements*. *Synthese* **140**<sup>1-2</sup> (2004), 97–119. <http://www.columbia.edu/~hg17/>; first publ. 2003.
- [28] *Oxford English Dictionary*, 2nd ed. Oxford University Press. (Oxford, 2009). First publ. 1857.
- [29] C. D. Broad: *Perception, Physics, and Reality; An Enquiry into the Information That Physical Science Can Supply about the Real* (Cambridge: Cambridge University Press, 1914). <https://archive.org/details/cu31924029026710>.
- [30] *Quantities and units*, 3rd ed. International Organization for Standardization. (Geneva, 1993).
- [31] *ANSI/IEEE Std 260.3-1993: American National Standard: Mathematical signs and symbols for use in physical sciences and technology*. Institute of Electrical and Electronics Engineers. (New York, 1993).
- [32] *Guide for the Use of the International System of Units (SI): NIST special publication 811, 1995 edition*. National Institute of Standards and Technology. (Washington, D.C., 1995). <http://physics.nist.gov/cuu/Uncertainty/bibliography.html>.
- [33] E. T. Jaynes: *Information theory and statistical mechanics*. *Phys. Rev.* **106**<sup>4</sup> (1957), 620–630. <http://bayes.wustl.edu/etj/node1.html>, see also [95].
- [34] E. T. Jaynes: *Information theory and statistical mechanics*. In: [96] (1963), pp. 181–218. <http://bayes.wustl.edu/etj/node1.html>.
- [35] I. J. Good: *Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables*. *Ann. Math. Stat.* **34**<sup>3</sup> (1963), 911–934.
- [36] E. T. Jaynes: *Foundations of probability theory and statistical mechanics*. In: [97] (1967), pp. 77–101. <http://bayes.wustl.edu/etj/node1.html>.

- [37] J. Aczél, Z. Daróczy: *On Measures of Information and Their Characterizations* (New York: Academic Press, 1975).
- [38] E. T. Jaynes: *Where do we Stand on Maximum Entropy?* In: [98] (1979), pp. 15–118. <http://bayes.wustl.edu/etj/node1.html>; repr. with an introduction in [99], pp. 210–314.
- [39] J. M. van Campenhout, T. M. Cover: *Maximum entropy and conditional probability*. IEEE Trans. Inform. Theor. **IT-27**<sup>4</sup> (1981), 483–489.
- [40] D. S. Sivia: *Bayesian inductive inference, maximum entropy, & neutron scattering*. Los Alamos Science **4**<sup>19</sup> (1990), 180–207. <http://www.fas.org/sgp/othergov/doe/lanl/pubs/number19.htm>.
- [41] S.-C. Fang, J. R. Rajasekera, H.-S. J. Tsao: *Entropy optimization and mathematical programming*, Reprint (New York: Springer, 1997).
- [42] G. L. Bretthorst: *The maximum entropy method of moments and Bayesian probability theory*. Am. Inst. Phys. Conf. Proc. **1553** (2013), 3–15. <http://bayes.wustl.edu/glb/BretthorstHistograms.pdf>.
- [43] R. B. Potts: *Note on the factorial moments of standard distributions*. Aust. J. Phys. **6**<sup>4</sup> (1953), 498–499.
- [44] S. Ross: *A First Course in Probability*, 8th ed. (Upper Saddle River, USA: Pearson, 2010). First publ. 1976.
- [45] W. Feller: *An Introduction to Probability Theory and Its Applications*. Vol. I, 3rd ed. (New York: John Wiley & Sons, 1968). First publ. 1950.
- [46] D. G. Kendall: *On finite and infinite sequences of exchangeable events*. Studia Sci. Math. Hung. **2** (1967), 319–327.
- [47] B. de Finetti: *Sulla proseguità di processi aleatori scambiabili*. Rend. Ist. Mat. Univ. Trieste **1** (1969), 53–67.
- [48] D. Heath, W. Sudderth: *De Finetti's theorem on exchangeable variables*. American Statistician **30**<sup>4</sup> (1976), 188–189.
- [49] P. Diaconis: *Finite forms of de Finetti's theorem on exchangeability*. Synthese **36**<sup>2</sup> (1977), 271–281. <http://statweb.stanford.edu/~sabatti/Stat370/synthese.pdf>.
- [50] P. Diaconis, D. Freedman: *Finite exchangeable sequences*. Ann. Prob. **8**<sup>4</sup> (1980), 745–764.
- [51] E. T. Jaynes: *Some applications and extensions of the de Finetti representation theorem*. In: *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti*. Ed. by P. K. Goel, A. Zellner (Amsterdam: North-Holland, 1986), p. 31. <http://bayes.wustl.edu/etj/node1.html>.

- [52] J. P. Burg: *Maximum entropy spectral analysis*. PhD thesis. (Stanford: Stanford University, 1975). <http://sepwww.stanford.edu/data/media/public/oldreports/sep06/>.
- [53] G. J. Erickson, C. R. Smith, eds.: *Maximum-Entropy and Bayesian Methods in Science and Engineering*. Vol. 2: *Applications* (Dordrecht: Kluwer Academic, 1988).
- [54] H. Robbins: *A remark on Stirling's formula*. *Am. Math. Monthly* **62**<sup>1</sup> (1955), 26–29.
- [55] Y. Weissman: *An improved analytical approximation to  $n!$* . *Am. J. Phys.* **51**<sup>1</sup> (1983), 9. See also [56–60].
- [56] N. D. Mermin: *Improving an improved analytical approximation to  $n!$* . *Am. J. Phys.* **51**<sup>9</sup> (1983), 776. See also [55; 57–60].
- [57] J. B. Dence: *Reply to a letter by Weissman on Stirling's approximation*. *Am. J. Phys.* **51**<sup>9</sup> (1983), 776–778. See also [55; 56; 58–60].
- [58] W. A. Bowers: *More about approximations to  $n!$* . *Am. J. Phys.* **51**<sup>9</sup> (1983), 778. See also [55–57; 59; 60].
- [59] J. D. Vedder: *Analytical approximations to  $n!$* . *Am. J. Phys.* **51**<sup>9</sup> (1983), 778–779. See also [55–58; 60].
- [60] R. K. Pathria: *Mathematical basis for Weissman's approximation to  $n!$* . *Am. J. Phys.* **53**<sup>1</sup> (1985), 81–82. See also [55–59].
- [61] P. G. L. Porta Mana: *On the relation between plausibility logic and the maximum-entropy principle: a numerical study*. (2009). arXiv:0911.2197.
- [62] J.-M. Bernardo, A. F. Smith: *Bayesian Theory* (Chichester: John Wiley & Sons, 1994).
- [63] S. Fortini, L. Ladelli, E. Regazzini: *Exchangeability, predictive distributions and parametric models*. *Sankhyā A* **62**<sup>1</sup> (2000), 86–109.
- [64] C. Maes, F. Redig, A. Van Moffaert: *The restriction of the Ising model to a layer*. *J. Stat. Phys.* **96**<sup>1</sup> (1999), 69–107. arXiv:math/9810094.
- [65] E. Schneidman et al.: *Weak pairwise correlations imply strongly correlated network states in a neural population*. *Nature* **440**<sup>7087</sup> (2006), 1007–1012. arXiv:q-bio/0512013, [http://www.weizmann.ac.il/neurobiology/labs/schneidman/The\\_Schneidman\\_Lab/Publications.html](http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html).
- [66] J. Shlens et al.: *The structure of multi-neuron firing patterns in primate retina*. *J. Neurosci.* **26**<sup>32</sup> (2006), 8254–8266. See also correction [100].
- [67] G. Tkačik et al.: *Ising models for networks of real neurons*. (2006). arXiv:q-bio/0611072.

- [68] O. Marre et al.: *Prediction of spatiotemporal patterns of neural activity from pairwise correlations*. Phys. Rev. Lett. **102**<sup>13</sup> (2009), 138101. arXiv:0903.0127.
- [69] G. Tkačik et al.: *Spin glass models for a network of real neurons*. (2009). arXiv: 0912.5409.
- [70] E. Ganmor, R. Segev, E. Schneidman: *Sparse low-order interaction network underlies a highly correlated and learnable neural population code*. Proc. Natl. Acad. Sci. (USA) **108**<sup>23</sup> (2011), 9679–9684. [http://www.weizmann.ac.il/neurobiology/labs/schneidman/The\\_Schneidman\\_Lab/Publications.html](http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html).
- [71] H. Shimazaki et al.: *State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data*. PLoS Computational Biology **8**<sup>3</sup> (2012), e1002385. <http://hdl.handle.net/1721.1/71752>.
- [72] G. Tkačik et al.: *The simplest maximum entropy model for collective behavior in a neural network*. J. Stat. Mech. **2013** (2013), P03011. arXiv:1207.6319.
- [73] H. Shimazaki et al.: *Simultaneous silence organizes structured higher-order interactions in neural populations*. Sci. Rep. **5** (2015), 9821.
- [74] V. Rostami et al. (2015). In preparation.
- [75] T. Hailperin: *Best possible inequalities for the probability of a logical function of events*. Am. Math. Monthly **72**<sup>4</sup> (1965), 343–359.
- [76] T. Hailperin: *Probability logic*. Notre Dame J. Formal Logic **25**<sup>3</sup> (1984), 198–212.
- [77] T. Hailperin: *Probability logic and combining evidence*. Hist. Philos. Logic **27**<sup>3</sup> (2006), 249–269.
- [78] H. Jeffreys: *Scientific Inference*, 3rd ed. (Cambridge: Cambridge University Press, 1973). First publ. 1931.
- [79] A. Gelman et al.: *Bayesian Data Analysis*, 2nd ed. (Boca Raton, USA: Chapman & Hall/CRC, 2004). First publ. 1995.
- [80] M. Ghosh, G. Meeden: *Bayesian Methods for Finite Population Sampling*, Reprint (Dordrecht: Springer, 1997).
- [81] O. Kallenberg: *Probabilistic Symmetries and Invariance Principles* (New York: Springer, 2005).
- [82] P. Gregory: *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support* (Cambridge: Cambridge University Press, 2005).
- [83] D. S. Sivia: *Data Analysis: A Bayesian Tutorial* (Oxford: Oxford University Press, 2006). Written with J. Skilling.
- [84] M. A. Ferreira, H. K. Lee: *Multiscale Modeling: A Bayesian Perspective* (New York: Springer, 2007).

- [85] A. P. Dawid: *Exchangeability and its ramifications*. In: [86] (2013). ch. 2, pp. 19–29.
- [86] P. Damien et al., eds.: *Bayesian Theory and Applications* (Oxford: Oxford University Press, 2013).
- [87] V. Rostami et al. (2015). In preparation.
- [88] W. A. Whitworth: *Choice and Chance: With One Thousand Exercises*, Repr. of 5th ed. (New York and London: Hafner, 1965). First publ. 1867.
- [89] W. A. Whitworth: *DCC Exercises: Including Hints for the Solution of All the Questions in Choice and Chance* (Cambridge: Deighton Bell & Co., 1897).
- [90] P. S. Laplace (Marquis de): *Œuvres complètes de Laplace. Tome septième : Théorie analytique des probabilités* (Paris: Gauthier-Villars, 1886). ‘Publiées sous les auspices de l’Académie des sciences, par MM. les secrétaires perpétuels’; <http://gallica.bnf.fr/notice?N=FRBNF30739022>.
- [91] F. P. Ramsey: *The Foundations of Mathematics: and other Logical Essays* (London: Routledge & Kegan Paul, 1950). Ed. by R. B. Braithwaite, first publ. 1931.
- [92] H. E. Kyburg Jr., H. E. Smokler, eds.: *Studies in Subjective Probability*, 2nd ed. (Huntington, USA: Robert E. Krieger, 1980). First publ. 1964.
- [93] L. Wittgenstein: *Philosophical Grammar*, Reprint (Oxford: Blackwell, 2004). Ed. by Rush Rhees, transl. by Anthony Kenny; written ca. 1933–34, first publ. 1974.
- [94] J. Hintikka, P. Suppes, eds.: *Aspects of Inductive Logic* (Amsterdam: North-Holland, 1966).
- [95] E. T. Jaynes: *Information theory and statistical mechanics. II*. Phys. Rev. **108**<sup>2</sup> (1957), 171–190. <http://bayes.wustl.edu/etj/node1.html>, see also [33].
- [96] K. W. Ford, ed.: *Statistical Physics* (New York: W. A. Benjamin, 1963).
- [97] M. Bunge, ed.: *Delaware Seminar in the Foundations of Physics* (Berlin: Springer, 1967).
- [98] R. D. Levine, M. Tribus, eds.: *The Maximum Entropy Formalism* (Cambridge, USA: The MIT Press, 1979).
- [99] E. T. Jaynes: *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Reprint (Dordrecht: Kluwer Academic, 1989). Ed. by R. D. Rosenkrantz. First publ. 1983.
- [100] J. Shlens et al.: *Correction, the structure of multi-neuron firing patterns in primate retina*. J. Neurosci. **28**<sup>5</sup> (2008), 1246. See [66].