

On the mathematical consequences of binning spike trains

Bruno Cessac, Arnaud Le Ny, Eva Löcherbach

► **To cite this version:**

Bruno Cessac, Arnaud Le Ny, Eva Löcherbach. On the mathematical consequences of binning spike trains. Neural Computation, Massachusetts Institute of Technology Press (MIT Press), 2017, 29 (1), pp.146-170. <http://www.mitpressjournals.org/doi/abs/10.1162/NECO_a_00898?journalCode=neco>. <hal-01351964>

HAL Id: hal-01351964

<https://hal.inria.fr/hal-01351964>

Submitted on 5 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the mathematical consequences of binning spike trains

B. Cessac*, A. Le Ny†, E. Löcherbach‡

August 5, 2016

Abstract

We initiate a mathematical analysis of hidden effects induced by binning spike trains of neurons. Assuming that the original spike train has been generated by a discrete Markov process, we show that binning generates a stochastic process which is *not Markov* any more, but is instead a Variable Length Markov Chain (VLMC) with unbounded memory. We also show that the law of the binned raster is a Gibbs measure in the DLR (Dobrushin-Lanford-Ruelle) sense coined in mathematical statistical mechanics. This allows the derivation of several important consequences on statistical properties of binned spike trains. In particular, we introduce the DLR framework as a natural setting to mathematically formalize anticipation, i.e. to tell "how good" our nervous system is at making predictions. In a probabilistic sense, this corresponds to condition a process by its future and we discuss how binning may affect our conclusions on this ability. We finally comment what could be the consequences of binning in the detection of spurious phase transitions or in the detection of wrong evidences of criticality.

1 Introduction

The development of multi-electrode arrays (MEA) technology offers an efficient way to record the spiking activity of populations of neurons, in the retina or in the cortex.

*Biovision team INRIA, Sophia Antipolis, France.

INRIA, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France.
email: bruno.cessac@inria.fr

†Laboratoire LAMA, UMR CNRS 8050, Créteil, France.

Université Paris Est Créteil (UPEC), 91 Avenue du Général de Gaulle, 94010 Créteil cedex, France.
email: arnaud.le-ny@u-pec.fr

‡Laboratoire AGM, UMR CNRS 8088, Cergy-Pontoise, France.

Université de Cergy-Pontoise, 2, avenue Adolphe Chauvin, 95302 Cergy-Pontoise Cedex, France.
email: eva.loecherbach@u-cergy.fr

Currently, up to 4096 neurons can be recorded simultaneously [12]. This provides new insights to better understand how a population of neurons encodes information.

The analysis of MEA data requires however preliminary specific treatments such as spike sorting, which allows to distinguish spikes coming from a specific neuron from the electric variations of potential recorded from nearby electrodes [5, 7, 21]. Once spikes have been sorted, and as spikes are sparse with a spike time subject to some indeterminacy, a usual strategy consists of binning data. That is, one defines first a time window of $\sim 5-20$ ms (binning window), larger than the typical duration of a spike ($\sim 1-2$ ms). The whole spike train is then divided into contiguous, non overlapping such windows and for each neuron a binary variable is defined: it takes value 0 if the neuron has not spiked in the binning window, and it is 1 if the neuron has spiked at least once; that is, windows containing 1, 2 or more spikes are all given the value 1.

Spike sorting, as well as binning, are operations which can have a strong impact on raw data, especially on spike train statistics. In this paper, we concentrate on the mathematical consequences of binning. Assuming that the original spike train has been generated by a Markov chain (i.e. a process with finite memory), we show that binning generates a process losing the Markov property. Instead, the stochastic process describing this raster is a Variable Length Markov Chain (VLMC) with unbounded memory. We will discuss in Section 4 that in some cases such a mechanism is expected to generate long range space and time correlations which might be misunderstood as fallacious evidence of phase transition or criticality. In our situation however, no phase transition and no phenomenon of criticality arises.

Anticipation plays a key role in the nervous system. A natural question is how "good" our nervous system is at making predictions. To assess the efficiency of prediction, a possible strategy is to measure the information that neurons carry about the future of sensory experiences. In this spirit, illuminating experiments have been done, in the retina, by S. Palmer *et al.* [23], who conclude that groups of neurons in the retina are indeed close to maximally efficient at separating predictive information from the non-predictive background. From a probabilistic point of view, such prediction amounts to condition a process by its future. In the context of mathematical statistical mechanics, this corresponds to the Dobrushin-Lanford-Ruelle (DLR) approach to rigorously define Gibbs measures and phase transitions [15]. DLR measures are measures consistent with regular systems of two-sided conditionings (i.e. conditionings w.r.t. the outside of finite sets), which corresponds to conditioning w.r.t. future and past in dimension one.

To see whether binning affects the capacity of prediction, we provide a first step towards this direction by proving that the Gibbs property in the DLR sense is preserved in our case. This means, loosely speaking, that the law of the binned raster behaves well in terms of capacity of prediction. We also discuss how one might get other situations where the Gibbs property is lost and what would be the consequences in the detection of spurious phase transitions or in the detection of wrong

evidences of criticality.

The paper is organized as follows. In Section 2 we give the main ideas explaining the effects announced above. Section 3 provides a rigorous setting for these statements. Finally, Section 4 is devoted to a discussion on the effect of binning in the context of MEA analysis. All mathematical proofs are given in the Appendix.

2 Qualitative description

2.1 Definitions

2.1.1 Spike trains

We consider the joint activity of N neurons, characterized by the emission of action potentials ("spikes"). The membrane potentials of neurons evolve according to known biophysical mechanisms [4, 16, 8]. Here, we consider that all what we are able to measure is spiking activity, e.g. via Multi-Electrode Arrays measurement followed by spike sorting.

We also assume that the spiking activity has been recorded at a time scale δ which is sufficiently small so that a neuron can at most fire one spike within a time window of size δ (we can set $\delta = 1$ without loss of generality). This provides a time discretization labeled with an integer time n . Each neuron's activity is then characterized by a binary variable $\omega_k(n) = 1$ if neuron k fires at time n and $\omega_k(n) = 0$ otherwise.

The state of the entire network at time n is thus described by a vector $\omega(n) \stackrel{\text{def}}{=} [\omega_k(n)]_{k=1}^N$. A spike block ω_m^n , $n \geq m$, is the sequence of vectors $\omega(m), \omega(m+1), \dots, \omega(n)$; blocks will be denoted by ω_m^n . The "time-range" (or "range") of a block ω_m^n is given by $n - m + 1$ which is the number of time steps needed to go from m to n . A "raster" (or "spike train") is a block $\omega_{n_0}^T$ where n_0 is the initial time of the experiment and T the final time. For convenience we shall often consider $n_0 \rightarrow -\infty$ and $T \rightarrow +\infty$, i.e. a bi-infinite raster. Thus, time index runs over \mathbb{Z} . For simplicity, we shall also use the notation ω for a raster.

2.1.2 Markov chain

We consider here the simple case where the spike train is a realization of a homogeneous Markov chain with memory $D > 0$. The evolution of the chain is characterized by a family of transition probabilities $\mathbb{P}[\omega(D) \mid \omega_0^{D-1}]$. "Homogeneous" means that these transition probabilities do not depend on time. To consider the simplest situation we assume that $\mathbb{P}[\omega(D) \mid \omega_0^{D-1}] > 0$ for all ω_0^D . Thus, the chain is *primitive*: there exists an n such that for any pair of blocs w', w of range D there is a path of length n and of positive probability joining w' to w .

An immediate consequence of primitivity is the existence and uniqueness of a unique invariant probability (equilibrium state). Additionally, time correlations decay exponentially fast [27, 2].

2.1.3 Binning

We fix an integer $\tau > 1$, the "binning window size". To a raster ω we associate a binned raster ϖ defined in the following way. We divide \mathbb{Z} into contiguous binning windows, $F_m = [m\tau, (m+1)\tau - 1] \cap \mathbb{Z}$, $m \in \mathbb{Z}$. We construct¹ the binned raster $\varpi(m) = [\varpi_k(m)]_{k=1}^N$

$$\varpi_k(m) = \begin{cases} 1, & \exists n \in F_m, \omega_k(n) = 1; \\ 0, & \forall n \in F_m, \omega_k(n) = 0. \end{cases} \quad (1)$$

Thus $\varpi_k(m) = 1$ if neuron k has spiked at least once in window F_m , and $\varpi_k(m) = 0$ if neuron k has never spiked in window F_m .

2.2 The consequences of binning

2.2.1 The simplest example

To start up, consider the case where $N = 1$ (one neuron), $D = 1$ (the Markov chain has memory one) and $\tau = 2$. This simple situation already captures the main features of the procedure. Thus, we can drop the neuron index 1 on the variable $\varpi_1(m)$ and write $\varpi(m)$.

As $\tau = 2$, the binned symbol $\varpi = 0$ corresponds to the successive events $(0, 0)$ in the initial raster. On the opposite, $\varpi = 1$ corresponds either to $(0, 1)$, $(1, 0)$ or $(1, 1)$. Thus, we associate to the symbol $\varpi = 1$ three symbols of the initial Markov chain. This operation is called "factorization". As we will show in Section 3, factorization leads in general to a loss of the Markov property with the creation of a *memory of variable length*. We illustrate this here. A general mathematical proof is given in the appendix.

We note \mathbb{P} the probabilities for the initial chain and $\mathbb{P}^{(b)}$ the probability for the binned chain. We want to show first how the binned chain loses the Markov property. For this we first show that

$$\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1, \varpi(0) = 1] \neq \mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1]. \quad (2)$$

¹This construction is similar to scaling transformations with non-overlapping blocks performed within the RG framework, such as decimation Kadanoff or majority-rule transformations. It is known in mathematical statistical mechanics that these transformations can generate hidden long-range orders due to untypical discontinuities, see e.g. [30, 29, 19, 32].

We have

$$\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1] = \frac{\mathbb{P}^{(b)}[\varpi(2) = 0, \varpi(1) = 1]}{\mathbb{P}^{(b)}[\varpi(1) = 1]}.$$

The complete formulation of $\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1]$ in terms of the transition probabilities $\mathbb{P}[\cdot | \cdot]$ of the initial Markov chain is easy using this formula and the Markov property, but the computation is relatively heavy, although simple. Thus, to make computations easier we use a diagrammatic expansion. An example is given in Fig. 1.

In order to compute $\mathbb{P}^{(b)}[\varpi(2) = 0, \varpi(1) = 1]$ in terms of the initial chain, one has to consider the symbols appearing on the line "P" of the diagram. Symbol $u|v$ corresponds to $\mathbb{P}[u | v]$, u to $\mathbb{P}[u]$, and we read the diagram from the left to the right, going from the present (on the left) to the past (on the right). One multiplies the probabilities on each path made with arrows, this gives a weight to this path. Finally, one adds all weights to obtain the probability. In this way we obtain (Fig. 1 left)

$$\mathbb{P}^{(b)}[\varpi(2) = 0, \varpi(1) = 1] = \begin{matrix} + & \mathbb{P}[\omega(5) = 0 | \omega(4) = 0] \mathbb{P}[\omega(4) = 0 | \omega(3) = 0] \mathbb{P}[\omega(3) = 0 | \omega(2) = 1] \mathbb{P}[\omega(2) = 1] \\ + & \mathbb{P}[\omega(5) = 0 | \omega(4) = 0] \mathbb{P}[\omega(4) = 0 | \omega(3) = 1] \mathbb{P}[\omega(3) = 1 | \omega(2) = 1] \mathbb{P}[\omega(2) = 1] \\ + & \mathbb{P}[\omega(5) = 0 | \omega(4) = 0] \mathbb{P}[\omega(4) = 0 | \omega(3) = 1] \mathbb{P}[\omega(3) = 1 | \omega(2) = 0] \mathbb{P}[\omega(2) = 0] \end{matrix}$$

and (Fig. 1 right)

$$\mathbb{P}^{(b)}[\varpi(1) = 1] = \begin{matrix} + & \mathbb{P}[\omega(3) = 0 | \omega(2) = 1] \mathbb{P}[\omega(2) = 1] \\ + & \mathbb{P}[\omega(3) = 1 | \omega(2) = 1] \mathbb{P}[\omega(2) = 1] \\ + & \mathbb{P}[\omega(3) = 1 | \omega(2) = 0] \mathbb{P}[\omega(2) = 0] \end{matrix}$$

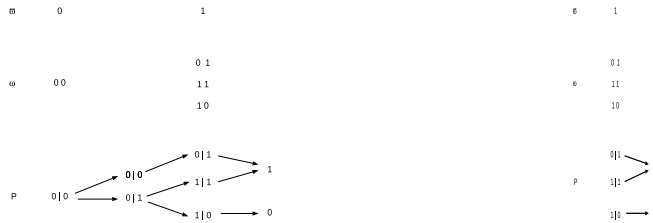


Figure 1: Left: $\mathbb{P}[\varpi(2) = 0, \varpi(1) = 1]$. Right: $\mathbb{P}[\varpi(1) = 1]$.

Finally, $\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1]$ is given by the ratio $\frac{\text{Left Diagram}}{\text{Right Diagram}}$, where by each "Diagram" we mean the product weight of admissible paths, see below.

As we have implicitly assumed stationarity here, we can in fact drop the time indexes as well, ending up with

$$\mathbb{P}^{(b)}[0|1] = \frac{\mathbb{P}[0|0] \mathbb{P}[0|0] \mathbb{P}[0|1] \mathbb{P}[1] + \mathbb{P}[0|0] \mathbb{P}[0|1] \mathbb{P}[1|1] \mathbb{P}[1] + \mathbb{P}[0|0] \mathbb{P}[0|1] \mathbb{P}[1|0] \mathbb{P}[0]}{\mathbb{P}[0|1] \mathbb{P}[1] + \mathbb{P}[1|1] \mathbb{P}[1] + \mathbb{P}[1|0] \mathbb{P}[0]}. \quad (3)$$

We shall use this shorthand notation from now on.

From the diagram in Fig. 2 we obtain $\mathbb{P}^{(b)}[0|11] = \frac{\mathbb{P}^{(b)}[011]}{\mathbb{P}^{(b)}[11]}$ as well.

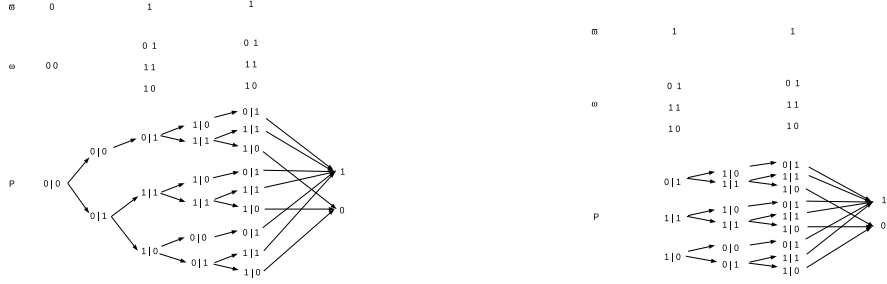


Figure 2: Left: $\mathbb{P}[011]$. Right: $\mathbb{P}[11]$.

Now, the reason why (2) holds can be readily seen on the graphs. The denominator of left hand side and right hand side are sums corresponding to paths either ending with $\mathbb{P}[1]$ or $\mathbb{P}[0]$. There is therefore no general way to simplify terms in the numerator and denominator so that the left-hand side equals the right-hand side.

Numerical example. We instantiate the above discussion by a concrete numerical example. We consider a Markov chain with transitions given by $\mathbb{P}[0|1] = \mathbb{P}[1|0] = \frac{3}{4}$, starting from $\omega(0) = 0$, and propose to calculate $\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1]$. Applying formula (3), we obtain that $\mathbb{P}^{(b)}[0|1] = \frac{5 \cdot 3^3 + 3}{4^2(3^3 + 7)} = \frac{138}{16 \cdot 34} \sim 0,2536$. On the other hand, we calculate in the same way $\mathbb{P}^{(b)}[0|10] = \frac{3^4}{4^2 \cdot 57} \sim 0,0888$ which is clearly different from $\mathbb{P}^{(b)}[0|1]$. Therefore,

$$\mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1, \varpi(0) = 0] \neq \mathbb{P}^{(b)}[\varpi(2) = 0 | \varpi(1) = 1],$$

which shows that the binned chain is not Markov of order one any more.

Let us come back to our general considerations. The situation is different when the conditioning term contains a “0”. Let us show that

$$\mathbb{P}^{(b)}[0|101] = \mathbb{P}^{(b)}[0|10]. \tag{4}$$

The probability $\mathbb{P}^{(b)}[0|101] = \frac{\mathbb{P}^{(b)}[0101]}{\mathbb{P}^{(b)}[101]}$ is represented in Fig. 3.

Applying the rules of our diagrams, we see that $\mathbb{P}^{(b)}[0101]$ and $\mathbb{P}^{(b)}[101]$ can be factorized into two subgraphs on the left and on the right of the central symbol 00. Thus the term on the right of these graphs disappears when computing the ratio

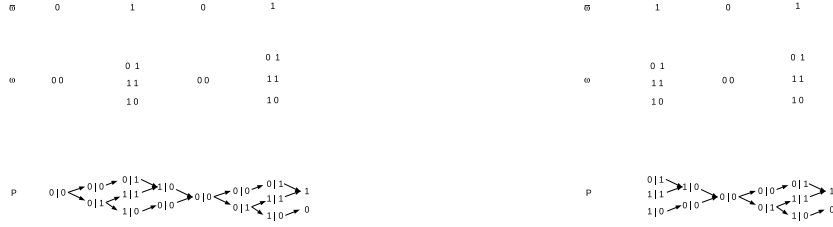


Figure 3: Left: $\mathbb{P}[0101]$. Right: $\mathbb{P}[101]$.

$\frac{\mathbb{P}^{(b)}[0101]}{\mathbb{P}^{(b)}[101]}$ and we end up with $\frac{\mathbb{P}^{(b)}[010]}{\mathbb{P}^{(b)}[10]} = \mathbb{P}^{(b)}[0|10]$.

We conclude now the above example. For a binned block ϖ_r^s we have

$$\mathbb{P}^{(b)}[\varpi(s+1)|\varpi_r^s] = \mathbb{P}^{(b)}[\varpi(s+1)|\varpi_l^s], \quad (5)$$

where l is the first occurrence of the symbol 0 when going from s to r (we set $l = r$, if ϖ does not contain the symbol 0). That is, the binned process is a *Variable Length Markov Chain* (VLMC) in the sense of Rissanen [26], where memory goes back up to the first occurrence of a 0 in the past (see also next the section where we formalise this idea).

2.2.2 Generalization

Let us now generalize this statement to N neurons (the rigorous proof is given in Section 3 and in the appendix). Consider first $\tau = 2$ and $N = 2$. We want to show e.g. that

$$\mathbb{P}^{(b)} \left[\begin{array}{c|c} 0 & 10 \\ \hline 0 & 01 \end{array} \right] \neq \mathbb{P}^{(b)} \left[\begin{array}{c|c} 0 & 1 \\ \hline 0 & 0 \end{array} \right].$$

The situation is basically the same as in the previous section. When computing $\mathbb{P}^{(b)} \left[\begin{array}{c} 010 \\ \hline 001 \end{array} \right]$ and $\mathbb{P}^{(b)} \left[\begin{array}{c} 10 \\ \hline 01 \end{array} \right]$, one has to construct a tree weighted by the transition probabilities of the initial chain.

To further generalize, we classify binned spikes patterns $\varpi(m)$ in two sets. The first set contains the unique pattern², denoted z , where all spikes in the window are 0. The second set contains all other patterns (at least one 1); all elements of this set are denoted by the symbol u . Now, (5) generalizes readily to this case, replacing

²From a general mathematical perspective the fact that z is a block of zeroes plays no specific role. The same argument works with z being any "end registry". We thank one of the referees for this remark.

"0" by "z" so that l becomes "the first occurrence of the symbol z when going from s to r ". We provide a schematic representation of this chain in Fig. 4.

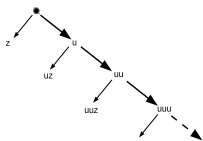


Figure 4: Schematic representation of the binned VLMC.

Let us now discuss the mathematical consequences of the fact that binning leads to a VLMC.

2.2.3 Consequences

The transformation of a Markov chain into a VLMC via binning has several consequences. We sketchy present them in this section. The mathematical justifications are given in Section 3.

1. The memory of the VLMC can extend arbitrarily far in the past. This is the case even if the initial raster is sparse. To obtain the symbol "u" one needs at least one spike in the binned window. In general, the probability of this event increases with τ , the binning window size, and N , the number of neurons.
2. The binned chain has therefore a long range memory, purely induced by binning.
3. This could induce fallacious long range time correlations as well as long range space correlations. Let us give a basic example. Assume that neurons are connected on a regular \mathbb{Z}^2 lattice with nearest neighbors interactions which are excitatory. Suppose neuron i_0 spikes and triggers a cascade of spikes (avalanche) spreading through the lattice. After $\tau + 1$ time steps the avalanche has reached neurons at distance $\tau + 1$ from i_0 . Binning will make those neurons and i_0 fire in contiguous time steps. This might create a fallacious causal interaction between them.

So, natural questions arise: "How much does binning impact the estimation of spikes statistics? How far can we mathematically control this impact ?" The next section is devoted to these questions.

3 Mathematical results

We write $\mathcal{A} = \{0, 1\}^N$ for the set of possible vectors with entries 0, 1 in a network of N neurons. The set of spike trains is $\Omega \equiv \mathcal{A}^{\mathbb{Z}}$. Moreover, for any fixed $n \in \mathbb{Z}$, we write $\mathcal{A}_{-\infty}^n$ for all infinite sequences $(\omega(k))_{-\infty < k \leq n}$, and we write $\omega_{-\infty}^n$ for short for such an infinite sequence.

We introduce the canonical random variable X_n on Ω defined as a projection by $X_n(\omega) = \omega(n)$ for all $n \in \mathbb{Z}$. We endow Ω with the product topology and the associated Borel σ -algebra \mathcal{F} which is generated by all projection maps X_n , i.e. $\mathcal{F} = \sigma\{X_n, -\infty < n < \infty\}$. Finally we introduce $\mathcal{F}_{-\infty}^k = \sigma\{X_n, -\infty < n \leq k\}$, the history before time k . Moreover, for any subset $\Lambda \subset \mathbb{Z}$, we introduce $\mathcal{F}_{\Lambda} = \sigma\{X_n, n \in \Lambda\}$.

A spike train $(\omega(n))_n$ can be seen as a stochastic process defined on the space (Ω, \mathcal{F}) , and its law is entirely determined by a probability law \mathbb{P} on (Ω, \mathcal{F}) .

3.1 Transition probabilities

In this paper, we want to analyze the effects of binning on inferring the present statistics of spike, given the history, but we also want to consider its effects on anticipation mechanism, where one conditions the present statistics by possible futures.

In a mathematical setting, this leads us to consider "one-sided" or "two-sided" dependencies. By "one-sided", we mean that the conditioning depends only on one side of the evolution – the past – and not on the other side – the future. By "two-sided", one means that the conditioning is prescribed outside finite sets from both "sides". In dimension one this amounts to considering a prescribed future. The two notions are not mathematically equivalent. In the simple case of range one dependencies, they are designed by *Local Markov* versus *Global Markov* properties, see [13, 17] or [10] in higher dimension. In this paper we are considering a one dimensional situation, where the dimension is time. But, more generally, considering one-sided or two-sided conditionings in larger dimensions is precisely the topic of rigorous mathematical statistical mechanics, see [30, 15]. These notions are also related to equilibrium states and Gibbs properties. We warn the interested reader that the vocabulary can change depending on the point of view which is adopted. Indeed, slight differences exist between the approaches coming from probability theory [15], mathematical statistical mechanics [6] or dynamical systems/ergodic theory [1]. See also [11] for a discussion on different notions.

For the sake of clarity and to avoid usual confusions, we precise here what we mean by Markov chains or fields, and which (different) Gibbs measures might be considered.

In the one-sided setting, we focus first on the probabilistic framework and consider stochastic processes $(X_n)_{n \in \mathbb{Z}}$ defined on (Ω, \mathcal{F}) . We follow [11] and introduce

systems of transition probabilities:

Definition 1 *A system of transition probabilities (or transition kernels) is a family $\{p_n(\cdot|\cdot) : n \in \mathbb{Z}\}$ of functions $p_n : \mathcal{A} \times \mathcal{A}_{-\infty}^n \rightarrow [0, 1]$, such that the following conditions hold for all $n \in \mathbb{Z}$.*

1. **Measurability** : *For each $a \in \mathcal{A}$, the function $p_n(a|\cdot)$ is $\mathcal{F}_{-\infty}^{n-1}$ -measurable.*
2. **Normalization** : *For each $\omega_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$,*

$$\sum_{a \in \mathcal{A}} p_n(a|\omega_{-\infty}^{n-1}) = 1.$$

A system of transition probabilities defines the intrinsic dynamics of the process $(X_n)_n : p_n(a|\omega_{-\infty}^{n-1})$ gives the probability of the event $\{X_n = a\}$, knowing that $X_{-\infty}^{n-1} = \omega_{-\infty}^{n-1}$. In other words, the law \mathbb{P} of $(X_n)_{n=-\infty}^{\infty}$ corresponds to the dynamics prescribed by $(p_n)_{n \in \mathbb{Z}}$. This is formalized in the following definition.

Definition 2 *A probability measure \mathbb{P} on (Ω, \mathcal{F}) is **consistent** with a system of transition probabilities $(p_n)_n$ if*

$$\mathbb{P}[X_n = \omega(n) | X_{-\infty}^{n-1} = \omega_{-\infty}^{n-1}] = p_n(\omega(n) | \omega_{-\infty}^{n-1}), \quad (6)$$

for all $n \in \mathbb{Z}$ and for \mathbb{P} -almost all $\omega_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$.

Here, $\mathbb{P}[X_n = \omega(n) | X_{-\infty}^{n-1} = \omega_{-\infty}^{n-1}]$ denotes a *regular version*³ of the conditional probability $\mathbb{P}[X_n = \omega(n) | \mathcal{F}_{-\infty}^{n-1}]$ on the event $\{X_k = \omega(k), -\infty < k \leq n-1\}$.

Let us emphasize that *stationarity is not assumed in these definitions and the kernel p_n may depend on n as well*. This means that seasonality can be included in the definition of the dynamics, although in the following we mainly assume stationarity.

Other related notions are those of *stochastic chains with memory of variable length* (VLMC, see below and [14]), chains with complete connections, g-measures or chains of infinite order (see e.g. [11], [31] and the references cited therein).

Homogeneous and primitive Markov chains are a particular case of such processes, see Definition 4 below. In the following, we assume stationarity so that it is sufficient to consider kernels p defined on $\mathcal{A} \times \mathcal{A}_{-\infty}^1$. Often, some additional continuity conditions (in the product topology of the discrete topology on our alphabet) are required that we recall now.

³This technical notion means the following: The transition kernel $p_n(\omega(n) | \omega_{-\infty}^{n-1})$ is one possible choice – within the L^2 -equivalence class of possible choices – ensuring that for all $A \in \mathcal{A}_{-\infty}^{n-1}$, $\mathbb{E}(1_{\{X_n = \omega(n)\}} 1_A) = \mathbb{E}(p_n(X_n | X_{-\infty}^{n-1}) 1_A)$. The notion *regular* does not refer to any regularity of the function $\omega_{-\infty}^{n-1} \mapsto p_n(\omega(n) | \omega_{-\infty}^{n-1})$ here, it is rather related to the fact that – very roughly speaking – it is possible to condition on an event of probability 0, that is, to condition on the event $\{X_k = \omega(k), -\infty < k \leq n-1\}$.

Definition 3 A transition kernel $p : \mathcal{A} \times \mathcal{A}_{-\infty}^{-1} \rightarrow [0, 1]$ is **continuous** if for all $a \in \mathcal{A}$ and for all $x_{-\infty}^{-1}$, $p(a|x_{-k}^{-1}y_{-\infty}^{-k-1})$ converges as $k \rightarrow \infty$, for any $y_{-\infty}^{-1}$. Here, $x_{-k}^{-1}y_{-\infty}^{-k-1}$ denotes the concatenated past given by the left-infinite sequence having element $y(l)$ for any $l \leq -k - 1$ and $x(l)$ for any $-k \leq l \leq -1$.

The continuity rate $\beta(k) \equiv \beta^p(k)$ of the kernel p is defined by

$$\beta(k) = \sup_a \sup_{x_{-k}^{-1}} \sup_{y_{-\infty}^{-1}, z_{-\infty}^{-1}} |p(a|x_{-k}^{-1}y_{-\infty}^{-k-1}) - p(a|x_{-k}^{-1}z_{-\infty}^{-k-1})|.$$

In other words, the dependency of the past decays with the distance between present and past, and this decay is described by the rate $\beta(k)$.

Markov chains are the simplest (non-independent) examples of processes consistent with a continuous kernel, because the conditioning depends on the immediate past only – more precisely, the immediate future X_{n+1} depends on the present X_n only. By extension, we call Markov chain of order D a stochastic chain having a transition kernel p which depends only of a finite portion, of length D , of the past. Such processes are trivial examples of processes having continuous transition kernels.

Definition 4 A system of transition kernels $p_n : \mathcal{A} \times \mathcal{A}_{-\infty}^{n-1} \rightarrow [0, 1]$ is a system of **Markov kernels of order $D > 0$** if for all $a \in \mathcal{A}, \omega \in \Omega$ and for all $n \in \mathbb{Z}$, $p_n(a|\omega_{-\infty}^{n-1})$ depends only on ω_{n-D}^{n-1} . A probability measure \mathbb{P} on (Ω, \mathcal{F}) is called **Markov measure of order D** if it is consistent with a system of transition probabilities $(p_n)_n$ which are Markov kernels of order D .

In this frame, a stochastic chain $(X_n)_{n \in \mathbb{Z}}$ canonically defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *Markov chain of order D* if and only if \mathbb{P} is a Markov measure of order D .

For homogeneous Markov chains of order D , i.e. chains where p_n does not depend on n , it suffices to study $p(\omega(0)|\omega_{-\infty}^{-1})$. Since this transition does only depend on ω_{-D}^{-1} , we shall write $p(\omega(0)|\omega_{-D}^{-1})$ instead of $p(\omega(0)|\omega_{-\infty}^{-1})$ in this case.

3.2 Modeling spike trains as Markov chains and binning

We model spike trains as realizations of a homogeneous and primitive Markov chain of order $D > 0$ having a transition kernel $p(\omega(0)|\omega_{-D}^{-1})$, where we suppose that $p(\omega(0)|\omega_{-D}^{-1}) > 0$ for all ω_{-D}^0 . We write \mathbb{P} for the unique Markov measure on (Ω, \mathcal{F}) which is consistent with this family of transition probabilities, where consistency has to be understood in the sense of Kolmogorov's consistency of marginals [2].

As in Section 2.1.3, we associate a binned raster $(\varpi(n))_n$ to a raster $(\omega(n))_n$. Moreover, we extend the notion of binning to binned blocks and write for any block $a_l^m \in \mathcal{A}_l^m$, $\varpi_l^m = a_l^m$ if all $\varpi(r) = a(r)$, $r = l, \dots, m$. We define a map $\pi : \mathcal{A}_0^{\tau-1} \rightarrow \mathcal{A}$, $\pi = [\pi_k]_{k=1}^N$, called *aggregation map*, by

$$\pi_k(\omega_0^{\tau-1}) = \begin{cases} 0, & \omega_k(n) = 0 \text{ for all } 0 \leq n \leq \tau - 1; \\ 1, & \text{else.} \end{cases} \quad (7)$$

π is extended in a canonical way to finite sequences $\omega_{l\tau}^{(m+1)\tau-1}$. As in Section 2.2.2, we write z for the element $[0]_{k=1}^N \in \mathcal{A}$. Thus, the set $\pi^{-1}(z_l^m)$ is the set of blocks $\omega_{l\tau}^{(m+1)\tau-1}$, in the original raster, such that each window F_r , $r = l, \dots, m$, contains only 0's. Finally, we denote by $\mathbb{P}^{(b)}$ the law of the binned raster, i.e. $\mathbb{P}^{(b)} = \mathcal{L}((\pi(X_{n\tau}^{(n+1)\tau-1}))_{n \in \mathbb{Z}} | \mathbb{P})$. We write $\mathbb{E}^{(b)}$ for the associated expectation whereas \mathbb{E} denotes expectation with respect to the original measure \mathbb{P} .

3.3 The binned raster is a Variable length memory chain (VLMC)

An immediate consequence of binning is that the resulting chain is not Markov any more, as we have argued in Section 2.2. Indeed, as we will show in this section, now rigorously and in a more general set-up, it is a chain of infinite memory having variable length memory. Stochastic chains with memory of variable length constitute an interesting family of stochastic chains of infinite order on a finite alphabet. The idea is that for each infinite past, a finite part of the past is enough to predict the next symbol, but the length of this past varies and can be arbitrarily long. These models were first introduced in the information theory literature by Rissanen [26] as a universal tool to perform data compression. For more details, we refer the reader to [14] for a survey of the subject.

In our framework, the variable length memory structure is given as follows. Write $\mathcal{A}_{-\infty}^{-1}$ for the set of all infinite pasts and define for any $x_{-\infty}^{-1} \in \mathcal{A}_{-\infty}^{-1}$,

$$l(x_{-\infty}^{-1}) = \inf \{m : x(-m) = z\}$$

where we recall that z denotes the "null" configuration for which no neuron has spiked. Here, by convention, $\inf \emptyset = -\infty$. Thus, $l(x_{-\infty}^{-1})$ is the first index, in the binned raster and back to the past starting from -1 , where the symbol z is met (i.e. the corresponding block contains no spike).

The following proposition shows that the memory of the chain is precisely $l(x_{-\infty}^{-1})$, as anticipated in Section 2.2.2. Thus, the length of the memory depends on the spike sequence.

Proposition 1 *Suppose that $\tau \geq D$. Then for any infinite past $x_{-\infty}^{-1}$ belonging to $\mathcal{A}_{-\infty}^{-1}$ and any symbol $a \in \mathcal{A}$,*

$$\mathbb{P}^{(b)} [X(0) = a | X_{-\infty}^{-1} = x_{-\infty}^{-1}] = \mathbb{P}^{(b)} \left[X(0) = a | X_{-l(x_{-\infty}^{-1})}^{-1} = x_{-l(x_{-\infty}^{-1})}^{-1} \right]. \quad (8)$$

Here, $\mathbb{P}^{(b)} [X(0) = a | X_{-\infty}^{-1} = x_{-\infty}^{-1}]$ denotes a version of the conditional probability $\mathbb{P}^{(b)} [X(0) = a | \mathcal{F}_{-\infty}^{-1}]$. The proof of this proposition is postponed to the appendix.

As a consequence, consider the tree \mathcal{T} represented in Figure 4 and defined by

$$\mathcal{T} = \{a_1^k z, k \geq 0, a(i) \in \mathcal{A}, a(i) \neq z, 1 \leq i \leq k\},$$

where $a_1^k z$ represents the sequence x_{-k-1}^{-1} such that $x(-i) = a(k-i+1)$ for all $1 \leq i \leq k$ and $x(-k-1) = z$. We associate transition probabilities to each leaf $a_1^k z$ of the tree via

$$p(a|a_1^k z) = \mathbb{P}^{(b)} [X(0) = a | X_{-1} = a_1, X_{-2} = a_2, \dots, X_{-k} = a_k, X_{-k-1} = z].$$

The ordered pair (\mathcal{T}, p) , where

$$p = \{p(\cdot | a_1^k z), k \geq 0\}$$

is called *probabilistic context tree on $\{0, 1\}^N$* . It defines entirely the evolution of the binned chain $(X_n)_{n \in \mathbb{Z}}$ on $(\Omega, \mathcal{F}, \mathbb{P}^{(b)})$ by (8).

Therefore, the binning procedure gives rise to a process which is a *Variable Length Markov Chain* (VLMC) where the memory extends to the last symbol z encountered in the past. Such a process is not Markov anymore, and we cannot bound a priori the memory depth of the chain. Indeed, this memory can go quite far back into the past. However,

$$\mathbb{P}^{(b)}(\exists \text{ infinitely many binning windows } F_m \text{ such that } X_m \equiv z) = 1,$$

since the original chain is primitive. Thus, with probability 1, the initial raster gives rise to a binned raster whose transition probabilities have memory depth which is finite (but not fixed, since the memory depends on the realization of the blocks).

3.4 Continuity properties of the binned transition operator

The transition kernel $p^{(b)}$ of the binned chain $(\varpi(n))_n$ is defined by:

$$\begin{aligned} p^{(b)} : \mathcal{A} \times \mathcal{A}_{-\infty}^{-1} &\rightarrow [0, 1] \\ (a, x_{-\infty}^{-1}) &\mapsto p^{(b)}[a|x_{-\infty}^{-1}] := \mathbb{P}^{(b)} [X_0 = a | X_{-\infty}^{-1} = x_{-\infty}^{-1}]. \end{aligned} \quad (9)$$

By Proposition 1, $p^{(b)}[a|x_{-\infty}^{-1}]$ depends only on $x_{-\ell}^{-1}$, if $l(x_{-\infty}^{-1}) = \ell$. But since $l(x_{-\infty}^{-1})$ is not bounded, $p^{(b)}$ is not of bounded memory.

We expect that the influence of past events on the probability $p^{(b)}(a|x_{-\infty}^{-1})$ of a decreases with their distance to this event: the further one goes back to the past, the less the past events influence the present. This question is related to the continuity properties of the kernel $p^{(b)}$. In general, this effect can decay either fast (e.g. exponentially) or slowly (e.g. sub-exponentially or algebraically). The consequences are quite different. In our situation, this decay is exponential as shows the following proposition.

Proposition 2 [Theorem 3.1 of [3]] *The transition kernel $p^{(b)}$ of the binned chain is continuous and there exists a constant $\alpha > 0$, depending on N , such that*

$$\beta^{p^{(b)}}(k) = O(e^{-\alpha k}) \text{ as } k \rightarrow \infty.$$

As a consequence of the above proposition, the transition operator of the binned chain is of infinite memory, but it is continuous and it has furthermore an exponential decay of the continuity rate. Note however that the decay rate, α , will in general depend on the number of neurons (see Section 4).

3.5 Does binning affect anticipation ?

In the context of VLMC's or more generally of chains of infinite order, exponential continuity is enough to ensure the existence and uniqueness of a probability measure consistent with the system of transition probabilities (see Section 3.1 and/or [9]). This is a sought property as it means that there is a unique invariant probability for the chain. However Proposition 2 states only the continuity of the one-sided transition probabilities obtained by conditioning upon the past. If we want to consider the effects of binning on anticipation, we might also want to consider conditioning on the future. In mathematical terms, this amounts of considering the Gibbs property in the DLR sense of mathematical statistical mechanics. In general, continuity with respect to the past does not imply that the law of the binned chain is also Gibbs in the DLR sense ([15, 6], see also [11] where an example is exhibited). Therefore, it is not a priori clear that the law of the binned raster is Gibbs in the DLR sense.

In the present section, we show that the law of the binned raster is a Gibbs measure in the DLR sense, i.e. that, roughly speaking, it behaves well when also conditioning with respect to the future. Here, *to behave well* means that the binned chain possesses the same good anticipation properties as the non-binned original chain and that a law of large numbers holds as well as good mixing properties. The main reason why this is so here is that the binned chain has very good regularity properties, i.e. the continuity rate is exponential ⁴as shown in Proposition 2.

Let us start by recalling the following definitions of Gibbs measures in the DLR sense from mathematical statistical mechanics (see e.g. [11]).

Definition 5 *A specification is a family of transition kernels $\gamma = \{\gamma_\Lambda\}_{\Lambda \subset \mathbb{Z}, |\Lambda| < \infty}$, $\gamma_\Lambda : \mathcal{F} \times \Omega \rightarrow [0, 1]$, on (Ω, \mathcal{F}) such that*

- (a) *For each $\Lambda \subset \mathbb{Z}$, $|\Lambda| < \infty$ and each $B \in \mathcal{F}$, the function $\gamma_\Lambda(B|\cdot)$ is \mathcal{F}_{Λ^c} -measurable.*
- (b) *For each $\Lambda \subset \mathbb{Z}$, $|\Lambda| < \infty$ and each $B \in \mathcal{F}_{\Lambda^c}$, $\gamma_\Lambda(B|\omega) = 1_B(\omega)$.*

⁴Note that this condition is not necessary.

(c) For any pair of regions Λ and Δ , with $\Lambda \subset \Delta \subset \mathbb{Z}$, $|\Delta| < \infty$, and any $B \in \mathcal{F}$,

$$\int_{\Omega} \gamma_{\Lambda}(B|\omega') \gamma_{\Delta}(d\omega'|\omega) = \gamma_{\Delta}(B|\omega) \quad (10)$$

for all $\omega \in \Omega$.

In our frame, we are mainly interested in *positive* specifications, i.e. $\gamma_{\Lambda}(B|\omega) > 0$ for all $B \neq \emptyset$, for all $\omega \in \Omega$. In this case, a specification is uniquely determined by the so-called *one-point specification* $\{\gamma_{\{i\}}(\cdot|\omega), i \in \mathbb{Z}, \omega \in \Omega\}$, see [15]. We write for short $\gamma_i(\omega(i)|\omega) = \gamma_{\{i\}}(\{\omega(i)\}|\omega)$. Intuitively, this is a candidate for the conditional law of X_i conditionally on $\{X_n = \omega(n), n \neq i\}$.

We can now introduce the notion of continuity for specifications as we did before for systems of transition probabilities, see again [11].

Definition 6 (a) A specification γ is called **continuous** if for all $i \in \mathbb{Z}$, $\gamma_i(\omega(i)|\cdot)$ is continuous for all $\omega(i) \in \mathcal{A}$, i.e.

$$\sup_{x, y \in \Omega: x_{-n}^m = y_{-n}^m} |\gamma_i(x(i)|x) - \gamma_i(y(i)|y)| \rightarrow 0$$

as $n, m \rightarrow \infty$.

(b) A specification γ is called **strongly non-null** if there exists a constant $c > 0$ such that for all $\omega(i) \in \mathcal{A}$, $\gamma_i(\omega(i)|\omega) \geq c > 0$.

Thus, being "continuous" means that at the same time the dependency on the past and on the future decays with their distance to the present. Condition (a) without (b) concerns so-called *quasilocal specifications* in mathematical statistical mechanics [15].

We now recall the notion of a Gibbs measure in the sense of mathematical statistical mechanics, i.e. in the DLR sense.

Definition 7 A shift invariant measure \mathbb{P} on (Ω, \mathcal{F}) is a **Gibbs measure** if it is consistent with a continuous and strongly non-null specification γ , i.e. for all $i \in \mathbb{Z}$,

$$\mathbb{P} [X_i = \omega(i) | X_{-\infty}^{i-1} = \omega_{-\infty}^{i-1}, X_{i+1}^{\infty} = \omega_{i+1}^{\infty}] = \gamma_i(\omega(i)|\omega)$$

for \mathbb{P} -almost all $\omega \in \Omega$.

Thus, roughly speaking, a Gibbs measure corresponds to a chain giving weight to every event (non-nullness) where the influence of both past and future on the current state decays with the distance.

Let us now return to the problem stated in the beginning of this section: Is there a unique probability measure compatible with a given past and a given future? In mathematical terms, this means: Is the invariant measure $\mathbb{P}^{(b)}$ of the binned *chain* (one-sided) a Gibbs measure in the DLR sense (two-sided)? The following theorem gives a positive answer.

Theorem 1 *The law $\mathbb{P}^{(b)}$ of the binned chain is a Gibbs measure.*

The proof of this theorem is given in the Appendix. It relies on the well-known Gibbsian character of processes having exponential continuity rate.

As a conclusion, the influence of past *and* future in estimating the probability to be in the present state decays exponentially with the distance, and there is a unique probability compatible with a given past and future.

4 Consequences and conclusion

In this paper we have considered rigorously mathematical effects of the binning procedure on the spike train analysis. Especially, we have shown that binning induces naturally long memory effects, even if the initial process is Markovian.

For a fixed system size, we have excluded possible spurious mathematical consequences of this artificial memory, such as a qualitatively different behavior of the binned chain from the behavior of the original, non-binned, chain. Indeed, when starting with a Markov chain, i.e. a process having finite memory, as a model for spike trains, then the binned chain, though of unbounded memory with variable length, will automatically present all good statistical features needed to study its longtime behavior. These good features are the renewal property, implying factorization of the past, and the exponential decay of the continuity rates. Here, "renewal property" means that the past can be cut into i.i.d. parts of pieces of history in between successive renewal events, implying the ergodic theorem as a simple consequence of the law of large numbers. In other words, for both, binned and original chain, we dispose of a law of large numbers and the convergence to equilibrium will be exponentially fast. Of course, statistical averages of functions will not be the same in the two processes, but their longtime behavior is of the same type. Moreover, although the Markov property is lost here in the one-sided situation, we prove that the Gibbs property – in the *DLR* sense – remains.

Notice that our proof holds only *when the number of neurons is finite*, and specific singularities in the binned chain, thoroughly analyzed in the context of mathematical statistical physics and coined in terms of *phase transitions*, could arise as $N \rightarrow \infty$. Indeed, the exponential decay coefficient α in Proposition 2 is positive. But it depends on N , and we cannot exclude that it converges to 0 as $N \rightarrow +\infty$. In this case, there might exist evidence of first order phase transition, even if the number of neurons is finite, because estimation of probabilities is based on finite rasters: taking a raster of length T with a given number of neurons N and extrapolating statistical properties of the underlying probability for an increasing number of neurons $n_0 < n_1 < \dots \leq N$ could lead to such effects.

A *second order phase transition*, associated to the notion of critical phenomena, corresponds to the situation where we have a unique probability compatible with

the specification, but where space and time correlations decay algebraically instead of exponentially. A small perturbation on one neuron could in this case trigger long range effects which are power law distributed. Critical phenomena are interesting because they can be classified according to a set of numbers called *critical exponents*. Remarkably, critical phenomena observed in nature can be classified into a few "universality classes" sharing the same set of critical exponents, see [18, 20]. There exist various methods to compute critical exponents, the most well known being the renormalization group analysis, [19, 32]. For this reason, researchers are actively seeking evidence of critical phenomena e.g. in the retina [28]. In our case, the binned system cannot exhibit a critical behavior for N finite. This is excluded by classical results on primitive Markov chains and the Perron-Frobenius theorem (spectral gap). However, again, one cannot exclude that binning induces spurious evidence of criticality as extrapolating with a growing number of neurons and we fear that, as $N \rightarrow \infty$, binning could dramatically change the value of the critical exponents, leading to wrong conclusions concerning the universality class.

Let us finally point out some mathematical directions of situations in which the mathematical consequences of the artificial memory spanned by the binning procedure could be much worse than in the situation described above. Suppose e.g. that the law of the original chain is given by a two-sided model exhibiting a phase transition⁵, like e.g. the long-range Dyson model with pair potentials that decay polynomially with parameter $1 < \alpha < 2$ (see Redig and Wang [25], van Enter *et al.* [29]). Then we suspect that the extra memory due to binning added to long range interactions could give rise to a non-Gibbsian measure as a consequence of the creation of a point of (two-sided) discontinuity. This discontinuity is NOT a critical phenomena but could be fallaciously interpreted as a manifestation of criticality; the discontinuity is just a proof of the fact that the measure is non-Gibbs, it does not correspond to a phase transition of any order. It might even be possible that – starting from a uniqueness measure of the Dyson-Ising specification – the binning procedure could yield a lower temperature long-range model for which a hidden phase transition occurs, and this could be misunderstood as the creation of criticality by binning.

Binning shares similarities with renormalization group transformations for which these types of pathologies (due to scaling transformations [30]) have also been detected, and explained as the manifestation of discontinuities of the renormalized (i.e. binned) process. These discontinuities could as well wrongly be interpreted as a critical phenomena. In other words, the mathematical question which is interesting in this context is the following. "Can binning induce fallacious evidences of phase transitions?". Heuristically, in statistical physics, a phase transition is observed in a system whose number of degrees of freedom (here: neurons) tends to infinity. On practical grounds, where neurons number is always finite, one proceeds

⁵Starting by an invariant measure for the two-sided model, not for the one-sided.

by considering increasing sizes and extrapolate to infinity. Efficient methods such as Finite-Size Scaling (see [24]) allow to nicely extrapolate the properties of the system in the infinite size limit (thermodynamic limit). So, in our case, one has to determine how binning could affect such an extrapolation.

From another point of view the effect of binning has been nicely discussed in a recent paper by I. Mastromatteo and M. Marsili [22]. They have shown that inference procedures used in statistical mechanics are likely to yield models which are close to a phase transition. Distinguishable models tend to accumulate close to critical points, where the susceptibility diverges in infinite systems in a region where the estimate of inferred parameters is most stable. Their paper suggests that spurious evidences of criticality can be inherent to the way data are considered. Our paper gives a similar warning, in a different context.

Acknowledgment. We would like to thank the referees for helpful and constructive criticism.

5 Appendix

Proof of Proposition 1 Fix $x_{-\infty}^{-1}$. If $l(x_{-\infty}^{-1}) = +\infty$, then (8) is trivially satisfied. Therefore, suppose that $l(x_{-\infty}^{-1}) = \ell < \infty$. Notice that this event is $\mathcal{F}_{-\ell}^{-1}$ -measurable. Let $A \in \mathcal{F}_{-\infty}^{-\ell-1}$. Denote by 1_A the indicatrix function of event A . Then by definition of $l(x_{-\infty}^{-1})$ and by definition of the binned raster by means of the aggregation map π ,

$$\mathbb{E}^{(b)}[1_a(X_0)1_A1_{X_{-\ell}^{-1}=x_{-\ell}^{-1}}] = \mathbb{E}[1_{\pi^{-1}(a)}(X_0^{\tau-1})1_{\pi^{-1}(A)}(X_{-\infty}^{-\ell\tau-1})1_{\pi^{-1}(x_{-\ell+1}^{-1})}1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z}]. \quad (11)$$

But the RHS can be written as

$$\begin{aligned} & \mathbb{E}[1_{\pi^{-1}(a)}(X_0^{\tau-1})1_{\pi^{-1}(A)}(X_{-\infty}^{-\ell\tau-1})1_{\pi^{-1}(x_{-\ell+1}^{-1})}1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z}] \\ &= \mathbb{E}\left[\mathbb{E}\left(1_{\pi^{-1}(a)}(X_0^{\tau-1})1_{\pi^{-1}(x_{-\ell+1}^{-1})}|\mathcal{F}_{-\infty}^{(-\ell+1)\tau-1}\right)1_{\pi^{-1}(A)}1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(1_{\pi^{-1}(a)}(X_0^{\tau-1})1_{\pi^{-1}(x_{-\ell+1}^{-1})}|X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z\right)1_{\pi^{-1}(A)}1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z}\right], \quad (12) \end{aligned}$$

where the second line holds trivially because the events $\pi^{-1}(A)$ and $\{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z\}$ are $\mathcal{F}_{-\infty}^{(-\ell+1)\tau-1}$ -measurable, whereas the last line follows from the Markov property of $(X_n)_n$ under \mathbb{P} of order D , the fact that $\tau \geq D$ and the fact that the block $X_{-\ell\tau}^{(-\ell+1)\tau-1}$ is constrained to be equal to z .

If we define

$$p(a|x_{-\ell}^{-1}) := \frac{\mathbb{E} \left(1_{\pi^{-1}(a)}(X_0^{\tau-1}) 1_{\pi^{-1}(x_{-\ell+1}^{-1})} | X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z \right)}{\mathbb{E} \left(1_{\pi^{-1}(x_{-\ell+1}^{-1})} | X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z \right)}, \quad (13)$$

we obtain, comparing to (11) and (12),

$$\begin{aligned} & \mathbb{E}^{(b)} [1_a(X_0) 1_A 1_{X_{-\ell}^{-1} = x_{-\ell}^{-1}}] \\ &= \mathbb{E} \left[p(a|x_{-\ell}^{-1}) \mathbb{E} \left(1_{\pi^{-1}(x_{-\ell+1}^{-1})} | X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z \right) 1_{\pi^{-1}(A)} 1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z} \right] \\ &= p(a|x_{-\ell}^{-1}) \mathbb{E} \left[\mathbb{E} \left(1_{\pi^{-1}(x_{-\ell+1}^{-1})} | X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z \right) 1_{\pi^{-1}(A)} 1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z} \right] \\ &= p(a|x_{-\ell}^{-1}) \mathbb{E} \left[1_{\pi^{-1}(A)} 1_{\pi^{-1}(x_{-\ell+1}^{-1})} 1_{X_{-\ell\tau}^{(-\ell+1)\tau-1} \equiv z} \right], \end{aligned}$$

whence

$$\mathbb{E}^{(b)} [1_a(X_0) 1_A 1_{X_{-\ell}^{-1} = x_{-\ell}^{-1}}] = p(a|x_{-\ell}^{-1}) \mathbb{E}^{(b)} [1_A 1_{X_{-\ell}^{-1} = x_{-\ell}^{-1}}],$$

and as a consequence, we have shown that $p(a|x_{-\ell}^{-1})$ is a regular version of the conditional probability of $\mathbb{P}^{(b)} [X(0) = a | \mathcal{F}_{-\infty}^{-1}]$ on $\{X_{-\infty}^{-1} = x_{-\infty}^{-1}\}$, when $l(x_{-\infty}^{-1}) = \ell$. Note that being of variable length, this system of transition kernels is by definition continuous on the set of all semi-infinite past sequences containing at least one z . Indeed, on this set, the transition kernels depend only on a finite portion of the past and are therefore, a fortiori, continuous. This concludes our proof. \bullet

Proof of Proposition 2 We suppose first that $\tau \geq D$. Introduce $Y_n := X_{n\tau}^{(n+1)\tau-1}$, $n \in \mathbb{Z}$. Since $\tau \geq D$, Y_n is a primitive and homogeneous Markov chain of order 1 under \mathbb{P} , and $\mathbb{P}^{(b)} = \mathcal{L}((\pi(Y_n))_{n \in \mathbb{Z}})$ is the law of the factor chain obtained through the factor map π introduced in (7). Write q for the transition kernel of Y . By our assumptions, q is a strictly positive continuous transition kernel which is locally constant, i.e.

$$\beta^q(k) = 0 \text{ for all } k \geq 2.$$

Then Theorem 1.1 of [31] can be applied, and it implies that $\beta^{p^{(b)}}(k) \rightarrow 0$ as $k \rightarrow \infty$, without however giving a precise rate of convergence.

In order to obtain a control on the rate of convergence, we rely on the results obtained in [3]. The notations used there are slightly different from ours, in particular, they work with right infinite sequences of symbols drawn from \mathcal{A} which represent all possible pasts. If we translate our objects into their framework, then $\mu := \mathcal{L}(X_{-\infty}^0 | \mathbb{P})$ is a D -step Markov measure in the sense of [3]. Then Equation (7) of Theorem 3.1 of [3] implies that there exists a function $\Phi^{(b)} : \mathcal{A}_{-\infty}^0 \rightarrow \mathbb{R}$ such that we have convergence

$$\lim_{n \rightarrow \infty} \log p^{(b)}(a|a_{-n}^{-1}) = \Phi^{(b)}(a_{-\infty}^0) \quad (14)$$

for all $a_{-\infty}^0 \in \mathcal{A}_{-\infty}^0$ and such that

$$\sup_{a_{-\infty}^0, b_{-\infty}^0 \in \mathcal{A}_{-\infty}^0 : a_{-n}^0 = b_{-n}^0} |\Phi^{(b)}(a_{-\infty}^0) - \Phi^{(b)}(b_{-\infty}^0)| \leq Ce^{-\alpha n}, \quad (15)$$

for some $\alpha > 0$. Due to the uniform continuity of $] - \infty, 0] \ni x \mapsto e^x$, this implies also that

$$\beta^{p^{(b)}}(n) \leq Ce^{-\alpha n},$$

for all n , and this concludes the proof.

Finally, if $\tau < D$, let $N = \min\{k : k\tau \geq D\}$ and set $\tilde{Y}_n = X_{n\tau}^{(n+N)\tau-1}$, $n \in \mathbb{Z}$. Then the above proof remains true, working with \tilde{Y}_n instead of Y_n . •

Proof of Theorem 1 Our proof follows ideas given in [3]. In [3], the authors prove the Gibbs property of a factor chain (i.e. of the law of the binned raster) in the sense of "Sinai-Ruelle-Bowen", see e.g. [1]. One speaks also shortly of "SRB"-Gibbs measures (cf. to Definition 2.3 and Theorem 3.1 of [3]). The SRB-Gibbs property is weaker than the standard DLR-Gibbs property in mathematical statistical mechanics and does not imply that conditioning with respect to the future behaves well. We refer to [11] for the hierarchy between the two notions.

The following proof shows that it is nevertheless possible to use the same approach as the one given in [3] to prove the Gibbs property also in the DLR sense.

By Theorem 2.8 of [11], it is sufficient to show the uniform convergence of

$$\mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}, X_1^n = a_1^n],$$

as $m, n \rightarrow \infty$. Under the conditions of our paper, this convergence follows easily from the considerations that we have developed in the proof of Proposition 2.

Indeed, let us rewrite

$$\mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}, X_1^n = a_1^n] = \frac{\mathbb{P}^{(b)} [X_{-m}^n = a_{-m}^n]}{\sum_{c \in \mathcal{A}} \mathbb{P}^{(b)} [X_{-m}^n = a_{-m}^{-1} c a_1^n]}. \quad (16)$$

We rewrite the numerator of this expression as

$$\begin{aligned} \mathbb{P}^{(b)} [X_{-m}^n = a_{-m}^n] &= \mathbb{P}^{(b)} [X_{-m}^{-1} = a_{-m}^{-1}] \\ &\quad \mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}] \mathbb{P}^{(b)} [X_1^n = a_1^n | X_{-m}^0 = a_{-m}^0]. \end{aligned}$$

The same kind of expression applies to the denominator. As a consequence,

$$\begin{aligned} \mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}, X_1^n = a_1^n] &= \\ &= \frac{\mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}] \mathbb{P}^{(b)} [X_1^n = a_1^n | X_{-m}^0 = a_{-m}^0]}{\sum_{c \in \mathcal{A}} \mathbb{P}^{(b)} [X_0 = c | X_{-m}^{-1} = a_{-m}^{-1}] \mathbb{P}^{(b)} [X_1^n = a_1^n | X_{-m}^0 = a_{-m}^{-1} c]}. \end{aligned} \quad (17)$$

But by (14), and following [3], we have uniform convergence of

$$\mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}] \rightarrow e^{\Phi^{(b)}(a_{-\infty}^0)}$$

and of

$$\mathbb{P}^{(b)} [X_0 = c | X_{-m}^{-1} = a_{-m}^{-1}] \rightarrow e^{\Phi^{(b)}(a_{-\infty}^{-1}c)},$$

as $m \rightarrow \infty$.

Moreover,

$$\mathbb{P}^{(b)} [X_1^n = a_1^n | X_{-m}^0 = a_{-m}^0] = \prod_{k=1}^n \mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}]$$

such that we can rewrite (17) as

$$\begin{aligned} \mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}, X_1^n = a_1^n] &= \mathbb{P}^{(b)} [X_0 = a(0) | X_{-m}^{-1} = a_{-m}^{-1}] \cdot \\ &\left[\sum_{c \in \mathcal{A}} \mathbb{P}^{(b)} [X_0 = c | X_{-m}^{-1} = a_{-m}^{-1}] \prod_{k=1}^n \left(\frac{\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}]}{\mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}]} \right) \right]^{-1} \end{aligned} \quad (18)$$

where $b(k) = a(k)$ for all $k \neq 0$, $b(0) = c$. It remains to show that

$$\prod_{k=1}^n \left(\frac{\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}]}{\mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}]} \right)$$

converges as $m, n \rightarrow \infty$. For any fixed k , we have, still by (14), that

$$\frac{\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}]}{\mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}]} \rightarrow e^{\Phi^{(b)}(a_{-\infty}^k) - \Phi^{(b)}(b_{-\infty}^k)} \quad \text{as } m \rightarrow \infty.$$

Moreover, we have that

$$\frac{\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}]}{\mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}]} \leq e^{C e^{-\alpha k}}, \quad (19)$$

which follows from the representation

$$\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}] = \int_{\Omega} p^{(b)}(b(k) | b_{-m}^{k-1} \omega_{-\infty}^{-m-1}) \mathbb{P}^{(b)}(d\omega) p^{(b)}(b(k) | b_{-m}^{k-1} \omega_{-\infty}^{-m-1}),$$

from (14) and from the fact that $b_1^k = a_1^k$.

Now, (19) implies by Lebesgue's theorem of dominated convergence that

$$\sum_{k=1}^n \log \left(\mathbb{P}^{(b)} [X_k = b(k) | X_{-m}^{k-1} = b_{-m}^{k-1}] \right) - \log \left(\mathbb{P}^{(b)} [X_k = a(k) | X_{-m}^{k-1} = a_{-m}^{k-1}] \right)$$

converges, as $n \rightarrow \infty$, and this concludes our proof. •

References

- [1] R. Bowen. Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Second revised version., volume 470 of Lect. Notes.in Math. Springer-Verlag, 2008.
- [2] P. Brémaud. Markov chains. Gibbs fields, Monte Carlo simulation, and queues. New York, NY: Springer, 1999.
- [3] J.R. Chazottes and E. Ugalde. On the preservation of Gibbsianness under amalgamation of symbols. Entropy of Hidden Markov Processes and Connections to Dynamical Systems, 81, 2011.
- [4] P. Dayan and L.F. Abbott. Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems. MIT Press, 2001.
- [5] M. Delescluse and C. Pouzat. Efficient spike-sorting of multi-state neurons using inter-spike intervals information. J Neurosci Methods, 150(1):16–29, 2006.
- [6] R.L. Dobrushin. The description of a random field by means of conditional probabilities and conditions of its regularity. Teor. Veroyatn. Primen., 13:201–229, 1968.
- [7] G. T. Einevoll, F. Franke, E. Hagen, C. Pouzat, and K. D. Harris. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. Current Opinion in Neurobiology, 22(1):11 – 17, 2012.
- [8] G. B. Ermentrout and D. H. Terman. Mathematical Foundations of Neuroscience. Springer, 1st edition. edition, July 2010.
- [9] R. Fernandez and G. Maillard. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. J. Stat. Phys., 118 (3-4):555–588, 2005.
- [10] R. Fernandez and C.-E. Pfister. Global specifications and non-quasilocality of projections of gibbs measures. Ann. Proba., 25(3):1284–315, 1997.
- [11] R. Fernandez, S. Gallo, and G. Maillard. Regular g -measures are not always Gibbsian. Electron. Commun. Probab., 16:732–740, 2011.
- [12] E. Ferrea, A. Maccione, L. Medrihan, T. Nieuw, D. Ghezzi, P. Baldelli, F. Benfenati, and L. Berdondini. Large-scale, high-resolution electrophysiological imaging of field potentials in brain slices with microelectronic multielectrode arrays. Frontiers in Neural Circuits., 6(80), 2012.

- [13] H. Föllmer. On the global Markov property. *Quantum fields - algebras, processes, Proc. Symp., Bielefeld 1978, 293-302 (1980).*, 1980.
- [14] A. Galves and E. Löcherbach. Stochastic chains with memory of variable length., pages 117–133. TICSP Series vol. 38, 2008.
- [15] H.O. Georgii. Gibbs Measures and Phase Transitions. De Gruyter Studies in Mathematics, Vol 9, Berlin New York, 1988.
- [16] W. Gerstner and W. Kistler. Spiking Neuron Models. Cambridge University Press, 2002.
- [17] S. Goldstein. Remarks on the global Markov property. Commun. Math. Phys., 74:223–234, 1980.
- [18] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. Rev. Mod. Phys., 49:435–479, Jul 1977. doi: 10.1103/RevModPhys.49.435.
- [19] L.P. Kadanoff. Scaling laws for Ising models near $T(c)$. Physics, 2:263–272, 1966.
- [20] Shang-keng Ma. Modern theory of critical phenomena. W. A. Benjamin, Advanced Book Program, 1976.
- [21] O. Marre, D. Amodei, N. Deshmukh, K. Sadeghi, F. Soo, T.E Holy, and M.J. Berry II. Mapping a Complete Neural Population in the Retina. The Journal of Neuroscience, 43(32):14859–14873, October 2012.
- [22] I. Mastro Matteo and M. Marsili. On the criticality of inferred models. J. Stat. Mech., page P10012, 2011.
- [23] S. E. Palmer, O. Marre, M. J. Berry II, and W. Bialek. Predictive information in a sensory population. PNAS, 112(22):6908–6913, 2015.
- [24] V. Privman and M. E. Fisher. Universal critical amplitudes in finite-size scaling. Phys. Rev. B, 30:322–327, Jul 1984.
- [25] F. Redig and F. Wang. Transformations of one-dimensional Gibbs measures with infinite range interaction. Markov Process. Relat. Fields, 16(4):737–752, 2010.
- [26] J. Rissanen. A universal data compression system. IEEE Trans. Inform. Theory, IT-29:656–664, 1983.
- [27] E. Seneta. Non-negative Matrices and Markov Chains. Springer, 2006.

- [28] G. Tkačik, Elad Schneidman, Michael J. Berry II, and William Bialek. Spin glass models for a network of real neurons. arXiv preprint arXiv:0912.5409, 2009.
- [29] A. van Enter and A. Le Ny. Decimation of the dyson-ising model. <http://arxiv.org/abs/1603.0549>, 2016.
- [30] A. C.D. van Enter, R. Fernández, and A. D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory. J. Stat. Phys., 72(5-6):879–1167, 1993.
- [31] E. Verbitskiy. On factors of g-measures. Indagationes Mathematicae, 22:315 – 329, 2011.
- [32] K.G. Wilson. The renormalization group: Critical phenomena and the kondo problem. Rev. Mod. Phys., 47:773–840, Oct 1975.