Stochastic models in neurobiology: from a multiunitary regime to EEG data

Aline Duarte de Oliveira

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Estatística
Orientador: Prof. Dr. Jefferson Antonio Galves

Durante o desenvolvimento deste trabalho a autora recebeu auxlio financeiro da CAPES e CNPq.
Esta tese foi produzida no Núcleo de Apoio à Pesquisa em Modelagem Estocástica e Complexidade/MaCLinC como parte das atividades do Centro de Pesquisa, Inovação e Difusão em Neuromatemática, financiado pela FAPESP (2013/07699-0)

São Paulo, julho de 2015

Stochastic models in Neurobiology: from the multiunitary regime to EEG data

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 17/07/2015. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Jefferson Antonio Galves (orientador) - IME-USP
- Prof. Dr. Pablo Augusto Ferrari - IME-USP
- Prof^a. Dr^a. Eva Löcherbach - U-Cergy
- Prof. Dr. Jacob Ricardo Fraiman Maus - U. San Andrés
- Prof^a. Dr^a. Claudia Domingues Vargas - UFRJ

Acknowledgments

Firstly, I would like to thank my parents, Degomar and Inês, who taught me the importance of education. They, together with my little sister Ariane, have given me plenty of support, love and comprehension and I will never be able to thank them enough.

I need to thank also Guilherme, a great partner in professional and personal life. During the last four years you was my best friend and my reference of safety and affection. After all this time growing together under lots of discoveries, challenges and hard working I am sure I could not have chosen a better companion for life.

I would like to express my gratitude to my supervisor, professor Antonio Galves, a talented teacher and enthusiastic researcher who taught me uncountable things. You have shown me the real spirit of science and collaboration and your leadership, fairness and hard work will be always a remarkable model to me. Moreover, I will be always grateful to you for the opportunity of being part of the great team NeuroMat.

I am in debt also to professor Errico Presutti for all clarifying discussions, for all the teaching, patience and hospitality during my visit in Gran Sasso Science Institute (GSSI).

I would like to thank NUMEC and all its goers (students, professors, cleaners and guards). It is a lab of excellence and at same time it seems as my second home. In particular, I would like to thank Lourdes for helping me so many times in so many subjects.

I also thank all friends I have made in São Paulo: Douglas Rodrigues, Rodrigo Lambert, Karina Yaginuma, Manuel González, Hugo Braga, Thiago Aloia, Aldana González, Bruno Monte, Andressa Cerqueira, Pedro Dias, Grabiel Peixoto, Michelle Miranda and many others. I have had an amazing time with you all.

Finally, I would to thank the department of statistics of the Universidade de São Paulo, the CAPES and CNPq for financial support and also the jury members of my PhD defense.

Abstract

In this thesis we study three different stochastic processes describing the brain activity. The first one is a continuous time version of the stochastic chains with memory of variable length. These stochastic chains take values in the set of neurons and assign, at time t, the value of the last neuron which spiked up to time t. Moreover, we assume neurons interact through a phenomena called chemical synapses. Briefly this means that when a neuron spikes, it loses all its membrane potential and at same time changes the membrane potential of the neurons which are influenced by it. Under this approach we proved the positive recurrent of the process and presented a perfect simulation algorithm able to generate a finite sample of the process under its invariant measure.

In the second model we continue considering the chemical synapses interaction and add also an interaction through electrical synapses. The last one happens duo to the presence of specific channels which allow the passage of ions along the the membrane of two neurons and, as consequence, we have a sharing of potential between the neurons. Moreover, we consider also the constant lost of potential of the neurons for the environment which push each neuron to a resting state. For this model we study the long-run behaviour of the process with a finite number of neurons, the hydrodynamic limit for the system and investigate the possible invariant distributions for the limiting process.

In the last model considered here we study the brain activity measured through EEG data. We investigate the predictive coding principle which says that neural networks are able to learn the statistical regularities inherent in a stimuli and reduce redundancy by removing the predictable components of the input. To test this conjecture we propose procedures to perform statistical model selection on the EEG data in order to retrieve structural features of stochastic sources. This is done through a case study in which the EEG data is recorded under the effect of two different stochastic rhythmic sources produced by two different context tree models. We present a suitable class of stochastic processes, called here as hidden context tree models, to model EEG signals evoked by rhythmic structures. Then, we propose a consistent statistical procedure to perform statistical model selection in this class and in our case study.

Key words : chains with memory of variable length, piecewise deterministic Markov process, limiting distribution, neuronal systems, hidden context tree models, statistical models selection

AMS Classification: 60K35, 60F99, 60J25

Resumo

Nessa tese estudamos três diferentes processos estocásticos descrevendo a atividade cerebral. O primeiro processo é uma versão a tempo contínuo das cadeias estocásticas com memória de alcance variável. Essas cadeias tomam valores no conjunto dos neurônios e assumem, no instante t, o valor do último neurônio a disparar antes de t. Além disso, assumimos que os neurônios interagem entre si através de fenômenos chamados sinapses químicas. Resumidamente isso significa que quando um neurônio dispara perde todo seu potencial de membrana e, simultaneamente, muda o potencial de membrana dos nerônios que influencia. Para esse processo estocásico provamos a recorrência positiva e apresentamos um algoritmo de simulação perfeita capaz de gerar uma amostra finita cuja ditribuição é a medida invariante do processo.

Na segunda classe de modelos continuamos considerando as sinápses químicas e adicionamos ainda interação por sinápses elétricas. A última acontece devido a presença de canais específicos entre dois neurônios que permitem a passagem de íons ao longo de suas membranas, como consequência, temos um compartilhamento de potencial entre os neurônios. Além disso, consideramos também a constante perda de potencial dos neurônios para o meio que age empurrando o potencial de cada neurônio a um estado de repouso. Com esses modelos estudamos o comportamento a longo prazo do processo com um número finito de neurônios, o limite hidrodinâmico desse sistema e investigamos a possível distribuição invariante para o processo limite.

Na última classe considerada aqui estudamos a atividade cerebral medida através de dados de EEG. Nós investigamos o princípio do código preditivo que afirma que redes neurais são capazes de aprender as regularidades estatísticas inerentes em um estímulo e reduzir a redundância removendo as componentes previsíveis. Para testar essa conjectura, propomos um procedimento para realizar seleção estatística de modelos em dados de EEG afim de recuperar características estruturais de fontes estocásticas. Isso é feito através de um caso de estudo em que dados de EEG são coletados sob o efeito de duas fontes ritímicas estocásticas distintas produzidas por duas árvores de contextos distintas. Nós apresentamos uma classe de modelos adequada, chamada aqui de modelos de árvore de contextos oculta, para modelar sinais de EEG evocados por estruturas rítmicas. Finalmente, propomos um procedimento estatístico consistente para fazer seleção estatística de modelos nessa nova classe assim como no nosso caso de estudo.

Contents

List of Figures

Chapter 1

Introduction

The brain is a complex network of a large number of components, the neurons. Recently much effort has been employed to describe the brain function through numerical simulation. Although this approach has been successfully applied for describing the behavior of single neurons, typically a single neuron interacts with more than 10^4 other neurons [GK02]. Thus, a simulation-based approach, although important, does not lead to a parsimonious understanding of the key mechanisms underlying the brain function. Even more, there are strong evidences suggesting that the brain both represents probability distributions and performs probabilistic inference [PBML13]. For this reason, in this thesis we address the neuronal activity in the stochastic processes framework.

We can model the brain processing in many levels, starting from the activity of a single neuron, taking into account the physiology of each part of the cell, going up to the synchronized activity of group of neurons (typically measured in EEG and fMRI techniques, for instance). In this thesis we study two different levels of the neuronal activity through three different classes of models.

Neurons communicate through electrical pulses, the *spikes*. These pulses, also called *action potentials*, are characterized by a fast and abrupt change of ions concentration in the interior of the cell. This change can be seen through the time evolution of the *membrane potential* which is the difference in voltage between the inside and outside of the cell membrane. An important fact is that spikes of a same neuron are highly similar, implying that the shape of an action potential does not improve the transfer of information [GK02]. For this reason, it makes sense to study the brain activity looking at *spike trains*, the sequence of spiking times of a neuron in a period of time. This is the motivation for the first two classes studied here.

In a microscopic level, we study the time evolution of the membrane potential of the neurons through a continuous time Markov chain which is also a example of piecewise deterministic Markov process. Still in the neurons level, we consider a continuous time version of the chains with memory of variable length which summarize the information of the spike trains of a set of neurons. Finally, we take into account the synchronized electrical activity of groups of neurons measured by *Electroencephalograph* (EEG) exams, and, in this scenario, we work with a stochastic model inspired by the analyses of experimental EEG data.

One of the earliest models in neurons was the *integrate-and-fire model* (IFM) [Lap07] used to reproduce spike trains. IFM captures the notion of the membrane potential being charged by currents flowing into it, firing an action potential and discharging, when the membrane potential exceeds a threshold ([SGGW11]). Classically, these models are used to describe the behavior of a single neuron inside a large network under the effect of a stochastic external input.

A more interesting approach of integrate-and-fire model was proposed by Galves and Löcherbach in [GL13]. They introduce a discrete-time class of models which is a non Markovian system of infinite interacting chains with memory of variable length. In their model a neuron does not spike when its membrane potential reach a threshold, but instead with a probability depending on its membrane potential. The authors proved the existence of the process by means of a perfect simulation algorithm and give an upper bound for the speed of the lost of memory of the process. A continuous time version of this class was introduced by De Masi et al. in [DMGLP15], where they

consider the spikes as stochastic point processes whose the spiking rate depend on the membrane potential. In [DMGLP15] it is proved that, as the number of neurons goes to infinity, the distribution of membrane potentials becomes deterministic and is described by a limit probability density which obeys a non-linear PDE of Hyperbolic type. This two paper inspired all models introduced here.

Our first model is a continuous time version of the stochastic chains with memory of variable length introduced by Rissanen [Ris83]. It is also a simplified continuous version of the Galves and Löcherbah model (GLM), for a finite set of neurons, and a particular case of the models in [DMGLP15]. But, instead of taking into account the set of spike trains or the evolution of membrane potentials we look at the time evolution of the spiking neurons. These stochastic chains take values in the set of neurons and assign, at time t, the value of the last neuron which spiked up to time t. Moreover, the interaction among neurons considered in these models happens through a phenomena called *chemical synapses*. Briefly this means that when a neuron spikes, it loses all its membrane potential and at same time changes the membrane potential of the neurons which are influenced by it. Under this approach we proved that this process is positive recurrent and present a perfect simulation algorithm able to generate a finite sample of the process under its invariant measure.

In the next class we continue considering the interaction between neurons through chemical synapses and add also interactions through *electrical synapses*. The last one happens when two neurons share a same channel between their membranes. This channel allows the passage of ions along of the two membranes and, as consequence, we have a sharing of potential between this neurons. In this models we take into account also a continuous interaction of the neurons with the environment. This interaction is due to *leak currents* which can be described as the constant lost of potential of the neurons for the environment, pushing each one of them to a resting state. Under all this interactions we study the long-run behavior of the process with a finite number of neurons, the hydrodynamic limit for this system and investigate the possible invariant distributions for the limiting process (the process we obtain by taking the number of neurons to infinity).

In the last part of this thesis we study a class of models inspired by the analysis of EEG data. We address the probabilistic inference problem in order to investigate the *predictive coding principle* which says the neural networks are able to learn the statistical regularities inherent in a stimulus and reduce the redundancy by removing the predictable components of the input. To test this conjecture we should be able to perform statistical model selection on EEG data and retrieve structural features of the stochastic sources used as stimuli during the experiment. Mathematically this implies to present a suitable class of stochastic processes, called here as *hidden variable length Markov chains*, to model EEG signals evoked by stochastic rhythmic structures. Then, we propose a consistent statistical procedure to perform statistical model selection in this class and apply the statistical method in a simulation study.

Chapter 2

Jumping Process with Memory of Variable Length

Although simple, the class of models studied in this chapter is a starting point for readers who are not familiar with Galves and Löcherbah [GL13] and De Masi et al.[DMGLP15] models. Moreover, this class helps to understand the intuition behind some proofs of the results presented in the Chapter 3.

Neuronal signals consist of short electrical pulses, the *action potentials* or *spikes*, and they are one of the ways neurons send information to each other. The brain activity can be represented through *spike trains* of the neurons which is a sequence of spiking times of a single neuron in a period of time [GK02]. Since we are working in continuous time, two neurons never spike together. This implies that, if we label the neurons, we can represent the set of all spike trains by the sequence of the spiking neurons. This is the intuition behind the class of models considered in this chapter.

We present a class of stochastic models describing the time evolution of a system of N neurons. These stochastic chains take values in the set of neurons and assign, at time t, the value of the last neuron which spiked up to time t.

The spiking times of a neuron are described here by point processes with rate depending on its *membrane potential*. The membrane potential of a neuron can be depicted as the accumulated activity of the system since its last spike. Besides, we consider that interactions among neurons happen only through *chemical synapses* which can be described in words as follows. When a neuron i spikes, its membrane potential is reset to the resting state, assumed here as 0. At the same time, this spike triggers a complex chain of biochemical processing resulting in a change of potential in the neurons which are influenced by i [GK02]. This reset of potential in the spiking neuron acts as a renewal event making the stochastic chains with memory of variable length good candidates to model a single neuron.

In this chapter, we assume a complete graph of interaction among the neurons, that is, each neuron influence and it is influenced by all other neurons. Moreover, we consider that all chemical synapses have the same weight (equal to 1). We proved the positive recurrence for the process and present a *perfect simulation* algorithm which can provide a finite sample with the same distribution of the invariant measure of the process.

In words the process evolves as follows. At time t, each neuron has a exponential clock $e_i, i = 1, \ldots, N$, whose intensity depends on its membrane potential. If the the clock associated to the neuron i rings before everything else (if $e_i = \min\{e_k, k = 1, \ldots, N\}$), then we say the neuron i spiked at time $t+e_i$, and our processes assume the value i at this instant. Afterwards, we update the membrane potential of all neurons, in consequence of the spike. That is, the membrane potential of i becomes 0 and the membrane potential of all the others is increased by 1. After this, we start again with new exponential clocks depending on the updated membrane potentials.

The class of stochastic chains considered here is a continuous time version of the stochastic chains with memory of variable length introduced by Rissanen [Ris83]. Our process is a simplified continuous version of the Galves and Löcherbah model, for a finite set of neurons, and a particular

case of the models presented in [DMGLP15].

2.1 Model definition and main results

Before defining the first model we introduce some notation and the definition of chain with memory of variable length which is crucial for the present chapter as well as in the Chapter 4.

Through this work A denote a finite alphabet and, given two integers numbers $-\infty < m \leq n < +\infty$, a_m^n denote the sequence $(a_m, \dots, a_n)$ of symbols in A and A^k the set of all sequences of symbols in A with length k. We write $x_{-\infty}^{-1} \in A^{-\mathbb{N}}$ for the semi-infinite sequences in $A^{(\dots,-n,\dots,-2,-1)}$. We use $|\cdot|$ to denote the number of elements of a set and $\ell(w_m^n) = n - m + 1$ to denote the length of a string.

Fixed two strings w and v of elements of A, we denote by vw the sequence in $A^{\ell(v)+\ell(w)}$ obtained by the concatenation of v and w. We say that a string $v = v_{-j}^{-1}$ is a *suffix* of w_{-k}^{-1} if $j \leq k$ and $v_{-i} = w_{-i}$ for all $i = 1, \dots, j$. This relation will be denoted by $v \preceq w$. If $j < k$ we say that v is a *proper suffix* and denote by $v \prec w$.

Definition 1. A finite subset τ of $A^* := \bigcup_{k=1}^{\infty} A^k$ is an *irreducible tree* if it satisfies the following conditions:

(i) *Suffix Property.* For no $w_{-k}^{-1} \in \tau$ we have $w_{-k+j}^{-1} \in \tau$ for $j = 1, \dots, k-1$.

(ii) *Irreducibility* No string belonging to τ can be replaced by a proper suffix without violating the suffix property.

We say that $\tau \preceq \tau'$ if for every $w' \in \tau'$ there exists $w \in \tau$ such that $w \preceq w'$. Naturally, as in the strings case, if $\tau \neq \tau'$ we shall write $\tau \prec \tau'$. Moreover, to simplify the presentation we will assume that all the context trees τ considered here are finite and we shall denote their height by $\ell(\tau) = \max\{\ell(w) : w \in \tau\}$.

Let $p = \{p(\cdot \mid w : w \in \tau\}$ be a family of probability measures on A indexed by the elements of τ. Each element of τ will be called *context* and we named the pair (τ, p) *probabilistic context tree*.

Definition 2. A stationary stochastic chain $(X_n)_{n\in\mathbb{Z}}$ taking values in A is called a *chain with memory of variable length* compatible with the probabilistic context tree (τ, p), if

(i) For any $m \geq \ell(\tau)$ and any sequence $x_{-m}^{-1} \in A^m$ with $\mathbb{P}(X_{-m}^{-1} = x_{-m}^{-1}) > 0$ it holds that

$$\mathbb{P}\big(X_0 = a \mid X_{-m}^{-1} = x_{-m}^{-1}\big) = p\big(a \mid c_\tau(x_{-m}^{-1})\big) \tag{2.1}$$

where $c_\tau(x_{-m}^{-1})$ is the only context in τ which is a suffix of x_{-m}^{-1}.

(ii) No proper suffix of $c_\tau(x_{-m}^{-1})$ satisfies (i).

Now we present the definition of the class of models we are interested in.

Definition 3. Let (τ, p) be a probabilistic context tree and $q : \tau \to (0, \infty)$ be a rate function associated to τ. We say that $(\eta(t))_{t\in\mathbb{R}}$ is *a jumping process with memory of variable length* associated to the probabilistic context tree (τ, p) and the rate function q, taking values in a finite set of $\mathcal{N}$, if the following conditions are fulfilled.

(i) There exists a sequence of random variables $(T_n)_{n\in\mathbb{Z}}$ satisfying

$$\dots < T_{-1} < T_0 \leq 0 < T_1 < T_2 < \dots$$

(ii) There exists a stochastic chain $(X_n)_{n\in\mathbb{Z}}$ with memory of variable length compatible with the probabilistic context tree (τ, p) such that

(a) $\mathbb{P}(T_{n+1} - T_n > t \mid X_{-\infty}^n) = e^{-tq(c_\tau(X_{-\infty}^n))}$ and

(b) $\eta(t) = X_n$ if $T_n \leq t < T_{n+1}$.

In this context we call the stochastic process $(X_n)_{n\in\mathbb{Z}}$ the *skeleton* of $(\eta(t))_{t\in\mathbb{R}}$.

In what follows we give one example of JPMVL, prove its positive recurrence and present a perfect simulation algorithm for it. Let $(X_n)_{n\in\mathbb{Z}}$ be a stochastic chain with memory of variable length, taking values in a finite set $\mathcal{N} = \{1, \ldots, N\}$, compatible with the probabilistic context tree τ which associates

$$\mathcal{N}^{(-\infty,\ldots,n)} \ni X_{-\infty}^n \mapsto c_\tau(X_{-\infty}^n) = X_{L_n}^n \in \tau, \tag{2.2}$$

where $L_n = \sup\{k \leq n; \{X_k, X_{k-1}, \ldots, X_n\} = \mathcal{N}\}$ is the shortest portion of the past in which each neuron spiked at least once, and with transition probabilities given by

$$p(i \mid X_{L_n}^n) = \frac{\varphi\big(\mathrm{U}_i(n)\big)}{\sum\limits_{j\in\mathcal{N}} \varphi\big(\mathrm{U}_j(n)\big)} \tag{2.3}$$

where $\varphi : \mathbb{N} \to [0, 1]$ is a non decreasing measurable function with $\varphi(0) = 0$ and $\varphi(u) > 0$ for any $u > 0$, $\mathrm{U}_i(n) = \sum_{j\neq i} \sum_{s=L_n^i}^n \mathbf{1}\{X_s = j\}$ is the accumulated activity of the system since the last spike of i and $L_n^i = \sup\{k \leq n; X_k = i\}$ is the last spiking time of the neuron i .

Consider a sequence of random variables $(T_n)_{n\in\mathbb{Z}}$ satisfying

$$\ldots < T_{-1} < T_0 \leq 0 < T_1 < T_2 < \ldots$$

and such that

$$\mathbb{P}(T_{n+1} - T_n > t \mid X_{-\infty}^n) = \exp\Big(- t \sum_{i\in\mathcal{N}} \varphi\big(\mathrm{U}_i(n)\big)\Big). \tag{2.4}$$

Now, if we define $\eta(t) = X_n$ when $T_n \leq t < T_{n+1}$, we have that $(\eta(t))_{t\in\mathbb{R}}$ is a jumping process with memory of variable length associated to the probabilistic context tree (τ, p) defined in (2.2) and (2.3), and the rate function $q : \tau \to (0, \infty)$ which associates

$$X_{L_n}^n \mapsto \sum_{i\in\mathcal{N}} \varphi\big(\mathrm{U}_i(n)\big). \tag{2.5}$$

In this scenario, we can think of $(\eta(t))_{t\in\mathbb{R}}$ as the time evolution of a finite set of neurons, where for each time t, we set $\eta(t) = i$, if $i \in \mathcal{N}$ was the last neuron which spiked before time t. Under this approach, the sequence $(T_n)_{n\in\mathbb{Z}}$ represents the spiking times of the system and $\mathrm{U}_i(n)$ describes the membrane potential of the neuron i at time $t \in [T_n, T_{n+1})$. Looking at (2.3), we see that the probability of a neuron i having a spike depends on the number of spikes of all other neurons in the system since the last spike of i. This fact implies that a neuron lost memory about the past when it spikes.

Our first result about the stochastic process $(\eta(t))_{t\in\mathbb{R}}$ is

Theorem 1. *The jumping process $(\eta(t))_{t\in\mathbb{R}}$ with memory of variable length associated to the probabilistic context tree (τ, p), satisfying (2.2) and (2.3), and with rate function $q : \tau \to (0, \infty)$ satisfying (2.5) is positive recurrent.*

By *perfect simulation* we mean generate a finite sample with the same distribution of the invariant measure of the process. The advantage of the perfect simulation is that the whole sample is generated from the invariant measure, we do not have to wait the convergence to it. This is the content of the next theorem.

Theorem 2. *$\mathbb{P}$-a.s. there exists an algorithm (presented in Section 2.1.2 below) which returns a sample of the unique stationary stochastic chain compatible with the jumping process with memory of variable length defined in the Theorem 1.*

2.1.1 The Positive Recurrence of $(\eta(t))_{t\in\mathbb{R}}$

The goal of this section is show that the process $(\eta(t))_{t\in\mathbb{R}}$ is positive recurrent. To this end, we shall look at the chain which describes the time evolution of membrane potentials of the neurons in $\mathcal{N}$.

Let $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ be the stochastic process taking values in the alphabet

$$\Omega := \{(u_1,\dots,u_N)\in\mathbb{N}^N : u_i\neq u_j,\ \forall i,j\in\mathcal{N},\ \text{and } \exists\ k\in\mathcal{N} \text{ s.t. } u_k=0\}$$

defined as $\mathrm{U}(n)=(\mathrm{U}_1(n),\dots,\mathrm{U}_N(n))$ where

$$\mathrm{U}_i(n)=\sum_{j\neq i}\sum_{s=L_n^i}^{n}\mathbf{1}\{X_s=j\}. \tag{2.6}$$

Note that, by the definition of $(\mathrm{U}(n))_{n\in\mathbb{Z}}$, when a neuron j spikes its membrane potential is reset and the membrane potential of all other neurons increase in one unit. This phenomenon can be described by the mapping $\Delta_i:\Omega\to\Omega$ defined as

$$\big(\Delta_i(u)\big)_j=\begin{cases} u_j+1, & \text{if } j\neq i\\ 0, & \text{if } j=i\end{cases}. \tag{2.7}$$

Observe that, $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ is a Markov chain whose transition probabilities are given by

$$p(\Delta_i(u)\mid u)=\frac{\varphi(u_i)}{\sum\limits_{j=1}^{N}\varphi(u_j)},\quad \forall i\in\mathcal{N}. \tag{2.8}$$

Example 2.1. Suppose that $X_{-5}^{7}=3514231352431$ is a sample of the chain with memory of variable length compatible with the probabilistic context tree (τ,p) given by (2.3) and (2.2) (see figure 2.1).

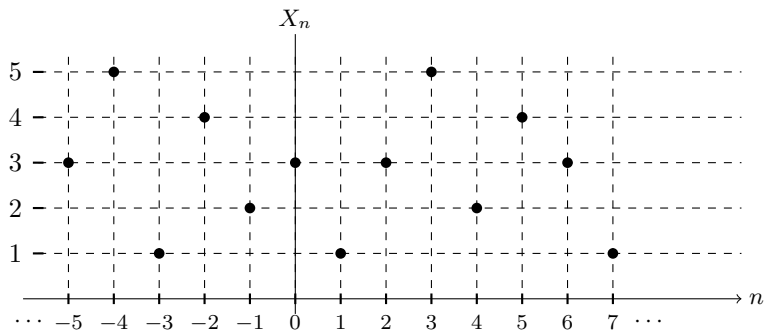


Figure 2.1: *Example of the time evolution of the process $(X_n)_{n\in\mathbb{Z}}$ for a system with 5 neurons*

In this case, the evolution of the membrane potentials of the chain is given by:

$$\begin{aligned}
\mathrm{U}_0&=(3,1,0,2,4) & \mathrm{U}_4&=(3,0,2,6,1)\\
\mathrm{U}_1&=(0,2,1,3,5) & \mathrm{U}_5&=(4,1,3,0,2)\\
\mathrm{U}_2&=(1,3,0,4,6) & \mathrm{U}_6&=(5,2,0,1,3)\\
\mathrm{U}_3&=(2,4,1,5,0) & \mathrm{U}_7&=(0,3,1,2,4)
\end{aligned}$$

By the definition of $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ in the equation (2.6), it is clear that there is a function $H^{X\to\mathrm{U}}:\tau\to\Omega$ which define the process $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ from the process $(X_n)_{n\in\mathbb{Z}}$. On the other hand, the

function $H^{\mathrm{U}\to X}:\Omega\to\mathcal{N}$ defined as

$$H^{\mathrm{U}\to X}(u)=\sum_{i\in\mathcal{N}} i\cdot\mathbf{1}\{u(i)=0\}, \tag{2.9}$$

describe the process $(X_n)_{n\in\mathbb{Z}}$ from the process $(\mathrm{U}(n))_{n\in\mathbb{Z}}$.

Although we can recover the information between the two processes without ambiguity both functions are not one-to-one. Indeed, it is not difficult to find examples of context which are different only in two coordinates, by a permutation between them, having the same potentials vector as image of $H^{X\to\mathrm{U}}$. For instance, both contexts $w=5421321$ and $\overline{w}=5412321$ have $(0,1,2,5,6)$ as correspondent vector of potentials. On the other hand, each vector of potentials with the i-th coordinate equal to 0 has i as image of the mapping $H^{\mathrm{U}\to X}$. Using this we can verify the next implication.

Proposition 2.1. *If the Markov chain $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ with transition probabilities given by (2.8) is positive recurrent, then the stochastic process $(X_n)_{n\in\mathbb{Z}}$ with memory of variable length compatible with the probabilistic context tree (τ,p) defined in (2.2) and (2.3) is positive recurrent.*

Proof. Suppose that $X_0=i$ and $\mathrm{U}(0)=u$ with $u_i=0$ and define the stopping times $T^{i\to i}=\inf\{n\geq 1: X_n=i\}$ and $T^{u\to u}=\inf\{n\geq 1:\mathrm{U}(n)=u\}$, the first time after zero that the processes $(X_n)_{n\in\mathbb{Z}}$ and $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ return to the state i and u, respectively. Now, observe that

$$\mathbb{E}_i[T^{i\to i}]=\sum_{k\geq 1}\mathbb{P}(T^{i\to i}=k\mid X(0)=i)\leq\sum_{k\geq 1}\mathbb{P}(T^{u\to u}=k\mid \mathrm{U}(0)=u)=\mathbb{E}_u[T^{u\to u}]<\infty$$

□

The proposition above implies that in order to prove the positive recurrence of $(X_n)_{n\in\mathbb{Z}}$ it is enough prove it for the process $(\mathrm{U}(n))_{n\in\mathbb{Z}}$. For this purpose, we will introduce another process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$. Then we will finally argue that the recurrence of $(X_n)_{n\in\mathbb{Z}}$ implies the recurrence of $(\eta(t))_{t\in\mathbb{R}}$.

Since $(\eta(t))_{t\in\mathbb{R}}$ is a continuous time process, two neurons never spike at the same time. Moreover, the fact that each spike of a neuron changes the membrane potential of all others by the same amount implies that two different neurons never have the same amount of potential. As a consequence, ordering the neurons by the time of its last spiking time is the same as to order them by the membrane potentials (from the maximum to the minimum). This rearrangement gives birth to the process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$.

Formally, consider the Markov process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$, taking values in

$$\tilde{\Omega}:=\{(\tilde{u}_1,\tilde{u}_2,\ldots,\tilde{u}_N)\in\mathbb{N}^N: 0=\tilde{u}_1<\tilde{u}_2<\ldots<\tilde{u}_N\},$$

defined, for each $n\in\mathbb{Z}$, by

$$\tilde{\mathrm{U}}(n)=\sigma_n(\mathrm{U}(n)):=\big(\mathrm{U}_{\sigma(1)}(n),\ldots,\mathrm{U}_{\sigma(N)}(n)\big), \tag{2.10}$$

where $\sigma_n:\mathcal{N}\to\mathcal{N}$ is the permutation such that $\mathrm{U}_{\sigma(1)}(n)<\mathrm{U}_{\sigma(2)}(n)<\ldots<\mathrm{U}_{\sigma(N)}(n)$.

Now, consider a sequence of random variables $(\tilde{I}_n)_{n\in\mathbb{Z}}$, taking values in the alphabet $\{1,\ldots,N\}$. In words, $\tilde{I}_n$ is the coordinate in $\tilde{\mathrm{U}}(n-1)$ of the neuron which spiked at time n, or equivalently, the coordinate in $\tilde{\mathrm{U}}(n-1)$ of the potential which was reset at time n. Mathematically,

$$\tilde{I}_n=i\iff\tilde{\mathrm{U}}_i(n-1)=\mathrm{U}_k(n-1)\text{ and }\mathrm{U}_k(n)=0. \tag{2.11}$$

To understand this definition, we give the following example

Example 2.2. Consider the sample X_{-5}^{7} of a chain with memory of variable length associated to the probabilistic context tree (τ,p) given in the example 2.1. In this case, by the definition of

$(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ and $(\tilde{I}_n)_{n\in\mathbb{Z}}$ we have:

$$\begin{array}{lcll} \mathrm{U}_4 = (3,0,4,6,1) & \xrightarrow{\sigma_4} & \tilde{\mathrm{U}}_4 = (0,1,2,3,6) & \\ \mathrm{U}_5 = (4,1,3,0,2) & \xrightarrow{\sigma_5} & \tilde{\mathrm{U}}_5 = (0,1,2,3,4) & \boxed{\tilde{I}_5 = 5} \\ \mathrm{U}_6 = (5,2,0,1,3) & \xrightarrow{\sigma_6} & \tilde{\mathrm{U}}_6 = (0,1,2,3,5) & \boxed{\tilde{I}_6 = 4} \\ \mathrm{U}_7 = (0,3,1,2,4) & \xrightarrow{\sigma_7} & \tilde{\mathrm{U}}_7 = (0,1,2,3,4). & \boxed{\tilde{I}_7 = 5} \end{array}$$

For instance, when $n = 7$, ordering by using the last spiking time implies $L_7^1 < L_7^3 < L_7^4 < L_7^2 < L_7^5$. With this, we can define the permutation $\sigma : \mathcal{I} \to \mathcal{I}$ given by $\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 4, \sigma(4) = 2$ and $\sigma(5) = 5$. This is described in the figure 2.2 below

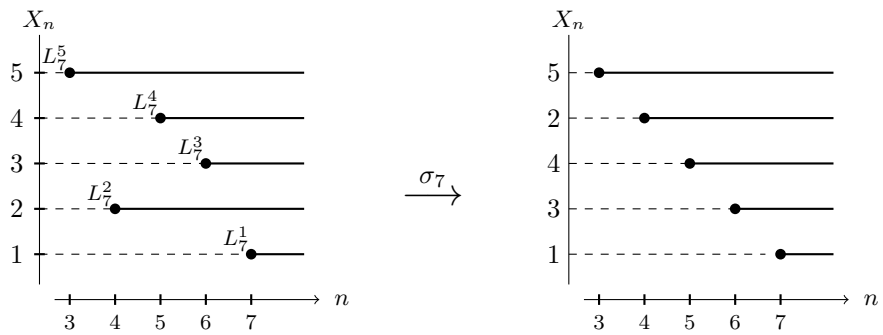


Figure 2.2: *Example of ordination by the last spiking time criterion.*

An important feature of the arranged process in that, for any non-null configuration of the process, if the process $\tilde{I}_n$ is equal to N in $N-1$ consecutive spiking times, then, in the $(N-1)$-th step, the process $\tilde{\mathrm{U}}_n$ will be in the state $(0,1,\ldots,N-1)$. Formally,

Lemma 2.1. *Given any symbol $u \in \Omega$, suppose that $\tilde{\mathrm{U}}(n) = \tilde{u}$. If $\tilde{I}_k = N$, for $k = n+1,\ldots,n+N-1$, then $\tilde{\mathrm{U}}(n+N-1) = (0,1,\ldots,N-1)$.*

Proof. Indeed, the event $tI_k = N$ for $k = n+1,\ldots,n+N-1$, implies the following sequence of implications

$$\begin{aligned} \tilde{\mathrm{U}}_n = (\tilde{u}_1,\tilde{u}_2,\tilde{u}_3,\tilde{u}_4,\ldots,\tilde{u}_N) \quad &\Rightarrow \quad \tilde{\mathrm{U}}_{n+1} = (0,\tilde{u}_1+1,\tilde{u}_2+1,\tilde{u}_3+1,\ldots,\tilde{u}_{N-1}+1) \\ &\Rightarrow \quad \tilde{\mathrm{U}}_{n+2} = (0,1,\tilde{u}_1+2,\tilde{u}_2+2,\ldots,\tilde{u}_{N-2}+2) \\ &\vdots \\ &\Rightarrow \quad \tilde{\mathrm{U}}_{n+N-1} = (0,1,2,3,\ldots,N-1). \end{aligned}$$

□

This lemma will be important in the prove of the following proposition.

Proposition 2.2. *The stochastic process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ defined in (2.10) is positive recurrent.*

Proof. Given any ordered configuration $\tilde{u} \in \tilde{\Omega}$, the probability of the neuron with the greatest potential having a spike at time n is

$$\mathbb{P}(\ \tilde{I}_n = N \mid \tilde{\mathrm{U}}(n-1) = \tilde{u}) = \frac{\varphi\left(\tilde{u}_{n-1}(N)\right)}{\sum_{i=1}^{N}\varphi\left(\tilde{u}_{n-1}(i)\right)} \geq \frac{\varphi\left(\tilde{u}_{n-1}(N)\right)}{N\varphi\left(\tilde{u}_{n-1}(N)\right)} = \frac{1}{N}, \tag{2.12}$$

where the inequality is justified by the fact of φ being a non decreasing function.

In order to conclude the proof we shall define a coupling between the random variables $(\tilde{I}_n)_{n\in\mathbb{Z}}$ and a random variable with geometric distribution of parameter $p = (1/N)^{N-1}$.

Consider the random variable K taking values in $\mathbb{N} \setminus \{0\}$ defined as

$$K = \inf \left\{ k \geq 1 : \bigcap_{i=(k-1)(N-1)+1}^{k(N-1)} \{\xi_i \leq 1/N\} \right\},$$

where $(\xi_n)_{n\in\mathbb{Z}}$ is a sequence of i.i.d random variables with uniform distribution in $[0,1)$. Under this assumptions, we have that K is a random variable with geometric distribution of parameter $p = (1/N)^{N-1}$, therefore $\mathbb{P}(K < \infty) = 1$.

We can coupling the random variable K with the sequence $(\tilde{I}_n)_{n\in\mathbb{Z}}$ on following way

(i) For each $\tilde{u} \in \tilde{\Omega}$ define a partition $(J_i^{\tilde{u}})_{i=1,\dots,N}$ of interval $[0,1)$ as

$$J_1^{\tilde{u}} = [0, p(N \mid \tilde{u})), \quad J_2^{\tilde{u}} = [p(N \mid \tilde{u}), p(N \mid \tilde{u}) + p(1 \mid \tilde{u})),$$
$$J_{k+1}^{\tilde{u}} = \left[p(N \mid \tilde{u}) + \sum_{i=1}^{k-1} p(i \mid \tilde{u}), p(N \mid \tilde{u}) + \sum_{i=1}^{k} p(i \mid \tilde{u}) \right); \ k = 3, \dots, N-1.$$

(ii) Define the function $G : \tilde{\Omega} \times [0,1) \to \{1, \dots, N\}$ by

$$G(\tilde{u}, z) = N \cdot \mathbf{1}\{J_1^{\tilde{u}}\}(z) + \sum_{i=2}^{N-1} (i-1)\mathbf{1}\{J_i^{\tilde{u}}\}(z);$$

(iii) Define for all $n \geq 1$, $\tilde{I}_n = G(\tilde{\mathrm{U}}(n-1), \xi_n)$.

Now, let $\tilde{T}^{(0,1,\dots,N)}$ be the the time of first return after zero of the process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ to the symbol $(0, 1, \dots, N)$. In this case,

$$\tilde{T}^{(0,1,\dots,N)} = \inf\{n \geq 1 : \tilde{\mathrm{U}}(n) = (0, 1, \dots, N)\}.$$

By the coupling above and the inequality (2.12) we have that

$$\{K = k\} \implies \bigcap_{i=(k-1)(N-1)+1}^{k(N-1)} \{\tilde{I}_i = N\},$$

and, by the remark 2.1, this implies that $\tilde{T}^{(0,1,\dots,N)} \leq K(N-1)$. □

Before prove the positive recurrence of $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ we need the following lemma,

Lemma 2.2. *Given $u \in \Omega$, define $T^{u\to u} = \inf\{n \geq 1 : \mathrm{U}(n) = u\}$. Then for all permutation $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ on the coordinates of u, it holds that*

$$\mathbb{E}(T^{u\to u} \mid \mathrm{U}(0) = u) = \mathbb{E}(T^{\sigma(u)\to\sigma(u)} \mid \mathrm{U}(0) = \sigma(u)).$$

Proof. Indeed for any $u, v \in \Omega$ and any permutation $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$, $p(u \mid v) = p(\sigma(u) \mid \sigma(v))$. Then for any $k \geq 1$,

$$\begin{aligned}
\mathbb{P}(T^{u\to u} = k \mid \mathrm{U}(0) = u) &= \sum_{u_1\dots,u_{k-1}:\ u_i \neq u} p(u_1 \mid u) \dots p(u \mid u_{k-1}) \\
&= \sum_{v_1\dots,v_{k-1}:\ v_i \neq \sigma(u)} p(v_1 \mid \sigma(u)) \dots p(\sigma(u) \mid v_{k-1}) \\
&= \mathbb{P}(T^{\sigma(u)\to\sigma(u)} = k \mid \mathrm{U}(0) = \sigma(u)). \qquad (2.13)
\end{aligned}$$

□

Analogously, for all $\tilde{u} \in \tilde{\Omega}$ define $\tilde{T}^{\tilde{u}\to\tilde{u}} = \inf\{n > 0 : \tilde{\mathrm{U}}(n) = \tilde{u}\}$ the time of first return after 0 of the process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ to the state $\tilde{u}$. It not difficult see that

Lemma 2.3. *For any vector $u \in \Omega$, define $\tilde{u} \in \tilde{\Omega}$ the ordered vector. The event when the process $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ return to u for the first time at same time that the process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ return to $\tilde{u}$ also for the first time has positive probability. Formally,*

$$\mathbb{P}(\tilde{T}^{\tilde{u}\to\tilde{u}} = T^{u\to u} \mid \mathrm{U}(0) = u) > 0.$$

Theorem 3. *The stochastic Markov chain $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ whose the transition probabilities are given by (2.8) is positive recurrent. In particular, we have that the stochastic chain $(X_n)_{n\in\mathbb{Z}}$ with memory of variable length compatible with the probabilistic context tree (τ, p), where τ is given by (2.2) and p by (2.3), is positive recurrent.*

Proof. First of all, observe that, since $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ is an irreducible and aperiodic Markov process, it is enough to prove that $\mathbb{E}(T^{u\to u} \mid \mathrm{U}(0) = u) < \infty$, for $u \in \Omega \cap \tilde{\Omega}$.

Let $u \in \Omega \cap \tilde{\Omega}$, $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ be a permutation and define $T^\sigma = \inf\{n \geq 1 : \mathrm{U}(n) = \sigma(u)\}$, then it holds that

$$\mathbb{E}(\tilde{T}^{u\to u} \mid \tilde{\mathrm{U}}(0) = u) = \sum_\sigma \mathbb{P}(\mathrm{U}(0) = \sigma(u) \mid \tilde{\mathrm{U}}(0) = u)\mathbb{E}(\tilde{T}^{u\to u} \mid \mathrm{U}(0) = \sigma(u)). \tag{2.14}$$

But, using that $\mathbb{P}(A \cap B) \leq \mathbb{P}(A)$ and the Lemma 2.2, we have that

$$\begin{aligned}\mathbb{E}(\tilde{T}^{u\to u} \mid \mathrm{U}(0) = \sigma(u)) &\geq \sum_{k\geq 1} k \cdot \mathbb{P}(\tilde{T}^{u\to u} = k, \tilde{T}^{u\to u} = T^\sigma \mid \mathrm{U}(0) = \sigma(u)) \\ &= \mathbb{P}(\tilde{T}^{u\to u} = T^\sigma \mid \mathrm{U}(0) = \sigma(u))\mathbb{E}(T^\sigma \mid \mathrm{U}(0) = \sigma(u))\end{aligned}$$

where in the equality use used that $\mathbb{P}(\tilde{T}^{u\to u} = k \mid \tilde{T}^{u\to u} = T^\sigma) = \mathbb{P}(T^\sigma = k)$. Finally, using Lemma 2.2 and replacing the last inequality in (2.14), we get

$$\begin{aligned}&\mathbb{E}(\tilde{T}^{u\to u} \mid \tilde{\mathrm{U}}(0) = u) \\ &\quad \geq \sum_\sigma \mathbb{P}(\mathrm{U}(0) = \sigma(u) \mid \tilde{\mathrm{U}}(0) = u)\mathbb{P}(\tilde{T}^{u\to u} = T^{u\to u} \mid \mathrm{U}(0) = u)\mathbb{E}(T^{u\to u} \mid \mathrm{U}(0) = u) \\ &\quad = \mathbb{P}(\tilde{T}^{u\to u} = T^{u\to u} \mid \mathrm{U}(0) = u)\mathbb{E}(T^{u\to u} \mid \mathrm{U}(0) = u).\end{aligned}$$

Then, the Lemma 2.3 implies that $\mathbb{E}(T^{u\to u} \mid \mathrm{U}(0) = u) < \infty$. □

In the next theorem we finally prove that the positive recurrence of the skeleton $(X_n)_{n\in\mathbb{Z}}$ implies the positive recurrence of the jumping process with memory of variable length $(\eta(t))_{t\in\mathbb{R}}$. For this end, we need introduce two sequences of stopping times $(T^{\eta,i})_{i\in\mathcal{N}}$ and $(T^{X,i})_{i\in\mathcal{N}}$ which are the first return time after zero of the processes $(\eta(t))_{t\in\mathbb{R}}$ and $(X_n)_{n\in\mathbb{Z}}$ to the symbol i. Formally,

$$T^{\eta,i} = \inf\{t > 0 : \eta(t) = i\} \qquad \text{and} \qquad T^{X,i} = \inf\{n \geq 1 : X_n = i\}. \tag{2.15}$$

Theorem 4. *Let $(\eta(t))_{t\in\mathbb{R}}$ be the jumping process with memory of variable length associated to the probabilistic context tree (τ, p) given by (2.2) and (2.3), and $(X_n)_{n\in\mathbb{Z}}$ be its skeleton . For any $i \in \mathcal{N}$, considering $T^{\eta,i}$ and $T^{X,i}$ as in (2.15), it holds that*

$$\mathbb{E}\left[T^{\eta,i} \mid \eta(0) = i\right] \leq \frac{1}{\sum_{l=0}^{N-1} \varphi(l)} \mathbb{E}\left[T^{X,i} \mid X(0) = i\right].$$

In particular, the stochastic jumping process $(\eta(t))_{t\in\mathbb{R}}$ is positive recurrent.

Proof. First of all, observe that for any time after each neuron having spiked at least once the minimum spiking rate of the system becomes $\sum_{l=1}^{N-1} \varphi(l) := q_0$. Moreover, the proof of Proposition

2.2 implies that the event each neuron spike at least once happens with probability 1. Therefore we shall assume, without lost of generality, that at time 0 the spiking rate of the system is at least q_0. Now, denoting $T^{\eta,i}$ and $T^{X,i}$, respectively by T^η and T^X, and $\mathbb{E}[A \mid \eta(0) = i] = \mathbb{E}_i[A]$ it holds,

$$\begin{aligned}\mathbb{E}_i[T^\eta] &= \sum_{k=1}^{\infty} \mathbb{E}_i[T^\eta = T_k \mid T_k < \infty]\mathbb{E}_i[T_k < \infty] \\ &\leq \frac{1}{q_0}\sum_{k=1}^{\infty} \mathbb{E}_i[\eta(T_k) = i \mid T_k < \infty]\end{aligned}$$

Seeing that $\mathbb{E}_i[\eta(T_k) = i \mid T_k < \infty] = \mathbb{E}[X(k) = i \mid X(0) = i]$, we finish the proof. □

Since the Markov chain $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ is positive recurrent suppose that μ is an invariant measure for process. The Lemma (2.2) and the Käc's Lemma implies that following result.

Lemma 2.4. *Let μ be an invariant measure of the chain of potentials $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ whose the transition probabilities are given by (2.8). For any $u \in \Omega$ and for any permutation σ of $\{1, \ldots, N\}$, it is truth that*

$$\mu(u) = \mu(\sigma(u)).$$

The next corollary states that, if we generate the processes $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ and $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$ in the same space of probabilities, given an ordered configuration $\tilde{u}$ of the process $(\tilde{\mathrm{U}}(n))_{n\in\mathbb{Z}}$, the probability of the non-ordered process being in a permutation of $\tilde{u}$ is uniform in the set of all permutations of $\{1, \ldots, N\}$.

Corollary 2.1. *Suppose that $\mathrm{U}(0) = u_0$, where u_0 was chosen using the invariant measure μ of the process $(\mathrm{U}(n))_{n\in\mathbb{Z}}$. Then, for any $n \geq 0$, conditioned on $\tilde{\mathrm{U}}(n) = \tilde{u}$, $\mathrm{U}(n)$ is chosen with distribution uniform in the set of all possible permutations of $\{1, \ldots, N\}$. In other words,*

$$\mathbb{P}(\mathrm{U}(n) = \sigma(u) \mid \tilde{\mathrm{U}}(n) = \tilde{u}) = 1/N!,$$

for any $n \geq 0$ and all σ permutation of $\{1, 2, \cdots, N\}$.

Proof. Indeed, for any $n \geq 0$,

$$\mathbb{P}(\mathrm{U}(n) = \sigma(u) \mid \tilde{\mathrm{U}}(n) = \tilde{u}) = \frac{\mathbb{P}(\mathrm{U}(n) = \sigma(u))}{\sum_{i=1}^{N!} \mathbb{P}(\mathrm{U}(n) = \sigma_i(u))} = \frac{\mu(\sigma(u))}{\sum_{i=1}^{N!} \mu(\sigma_i(u))} = \frac{1}{N!},$$

where the last equality comes from the lemma (2.4). □

This corollary give us an idea of how we could to do a perfect simulation of the process $(\eta(t))_{t\in\mathbb{R}}$.

2.1.2 Perfect Simulation

The goal of this section is present an algorithm able to generate, for any $-\infty < m < n < \infty$, a sample $X_m, \ldots, X_n$ of the stationary process $(\eta(t))_{t\in\mathbb{R}}$.

Let $(\eta(t))_{t\in\mathbb{R}}$ be a jumping process with memory of variable length associated to the probabilistic context tree (τ, p) defined in (2.2) and (2.3) and the rate function q described in (2.5). Fixed the two constants $-\infty < m < n < \infty$, consider a sequence of i.i.d random variables $(\xi)_{n\in\mathbb{Z}}$ and define the event

$$\theta[m, n] = \sup\left\{k \leq m - N \ : \ \bigcap_{l=k}^{k+N-1} \{\xi_l < 1/N\}\right\}. \tag{2.16}$$

Observe that in the Proposition 2.2 we proved that $\mathbb{P}(\theta[m, n] < -\infty) = 1$. Moreover,

$$\begin{aligned}\theta[m, n] &\implies \tilde{I}_{\theta[m,n]} = \ldots = \tilde{I}_{\theta[m,n]+N-1} = N \\ &\implies \{X(\theta[m, n]), \ldots, X(\theta[m, n] + N - 1)\} = \mathcal{N},\end{aligned}$$

where the sequence $(\tilde{I}_n)_{n\in\mathbb{Z}}$ is the one defined in (2.11), and this implies that the process $(X_n)_{n\in\mathbb{Z}}$ loses memory at $\theta[m,n]$.

Define, for each $w \in \tau$, the following partition of the interval $[0,1)$

$$J^X(1 \mid w) = \big[0, p(1 \mid w)\big) \text{ and } J^X(i \mid w) = \Big[\sum_{k=1}^{i-1} p(k \mid w), \sum_{k=1}^{i} p(k \mid w)\Big), \tag{2.17}$$

for $i = 2, \ldots, N$. By construction, $\lambda(J^X(i \mid w)) = p(i \mid w)$, where λ denote the Lebesgue measure in the the interval $[0,1)$.

In order to construct of the jumping process $(\eta(t))_{t\in\mathbb{R}}$ we need define two functions. The first one is a function $F : [0,1) \times \mathcal{N}^{-\mathbb{N}} \to I$ which give us a time evolution of the skeleton of $(X_n)_{n\in\mathbb{Z}}$, given whole the past

$$F(z, x_{-\infty}^{-1}) = \sum_{i\in\mathcal{N}} i \cdot \mathbf{1}\{z \in J^X(i \mid c_\tau(x_{-\infty}^{-1})\}. \tag{2.18}$$

Note that, if we define $X_{n+1} = F(\xi_{n+1}, X_{-\infty}^n)$, then

$$\mathbb{P}(F(\xi_{n+1}, X_{-\infty}^n) = i) = \mathbb{P}(\xi_{n+1} \in J^X(i \mid c_\tau(X_{-\infty}^n)) = p(i \mid c_\tau(X_{-\infty}^n)). \tag{2.19}$$

So, we can generate a sample of the chain $(X_n)_{n\in\mathbb{Z}}$ using the function f and the partitions J^X as in (2.18).

Now, consider a second function $G : [0,1) \times \mathcal{N}^{-\mathbb{N}} \times \mathbb{R} \to \mathcal{N} \times \mathbb{R}$ defined as

$$G(z, x_{-\infty}^n, t) = \Big(F(z, x_{-\infty}^n)\ ,\ t - \frac{1}{q(c_\tau(x_{-\infty}^n))} \log z\Big). \tag{2.20}$$

Then, the algorithm is given by

Theorem 5 (Existence and uniqueness)**.** *$\mathbb{P}$-a.s for all $-\infty < m \le n < +\infty$, the Algorithm **??** returns a sample of the unique stationary chain compatible with the jumping process $(\eta(t))_{t\in\mathbb{R}}$ with memory of variable length associated to the probabilistic context tree (τ, p), satisfying (2.2) and (2.3), and with rate function $q : \tau \to (0,\infty)$ satisfying (2.5).*

Compatibility proof. Since $\theta[m,n]$ is finite and by definition of τ, for all $n \in \mathbb{Z}$, always holds that $c_\tau(X_{-\infty}^{n+1}) \preceq c_\tau(X_{-\infty}^n)X_{n+1}$, this implies that $\theta[m+1,n] \le \theta[m,n]$, for all $m \le n$. Therefore, the sample which the Algorithm **??** returns

$$\big\{(X(\xi_{\theta[m,n]}), T(\xi_{\theta[m,n]})), (X(\xi_{\theta[m,n]+1}), T(\xi_{\theta[m,n]+1})), \ldots, (X(\xi_n), T(\xi_n))\big\}$$

is the same sample which we obtain using the whole past, and we can define it recursively for $\theta[m,n] \le j \le n$ by using that

$$(X_j, T_j) = \Big(F(\xi_j, X_{\theta[m,n]}^{j-1})\ ,\ T_{j-1} - \frac{1}{q(c_\tau(X_{\theta[m,n]}^{j-1}))} \log \xi_j\Big). \tag{2.21}$$

We need to show $(X(\xi_n))_{n\in\mathbb{Z}}$ is a chain with memory of variable length compatible with the probability tree (τ, p), as in the assumption of theorem, and also the sequence of random variables $(T_n(\xi_n) - T_{n-1}(\xi_{n-1}))_{n\in\mathbb{Z}}$ have exponential distribution with rate $q(c_\tau(X_{-\infty}^n))$, when $X_{-\infty}^n$ is given.

First of all, observe that, since the random variables $(\xi_n)_{n\in\mathbb{Z}}$ are independent

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{X(\xi_0) = i\} \mid \xi_{-\infty}^{-1}] &= \mathbb{E}[\mathbf{1}\{\xi_0 \in J^X(i \mid c_\tau(X(\xi_{-\infty}^{-1})))\} \mid \xi_{-\infty}^{-1}] \\
&= \mathbb{E}[\mathbf{1}\{\xi_0 \in J^X(i \mid c_\tau(X(\xi_{-\infty}^{-1})))\}] \\
&= p(i \mid c_\tau(X(\xi_{-\infty}^{-1}))),
\end{aligned}$$

what implies that $\mathbb{E}[\mathbf{1}\{X(\xi_0) = i\} \mid \xi_{-\infty}^{-1}]$ is $\mathcal{F}(X(\xi_{-\infty}^{-1}))$-measurable. Thus, since $\mathcal{F}(X(\xi_{-\infty}^{-1})) \subset \mathcal{F}(\xi_{-\infty}^{-1})$, we get that

$$\mathbb{E}\left[\mathbf{1}\{X(\xi_0) = i\} \mid X(\xi_{-\infty}^{-1})\right] = \mathbb{E}\left[\mathbb{E}[\mathbf{1}\{X(\xi_0) = i\} \mid \xi_{-\infty}^{-1}] \mid X(\xi_{-\infty}^{-1})\right] = p(i \mid c_\tau(X(\xi_{-\infty}^{-1}))).$$

In other words, $\mathbb{P}(X(\xi_0) = i \mid X(\xi_{-\infty}^{-1})) = p(i \mid c_\tau(X(\xi_{-\infty}^{-1})))$ and $(X(\xi_n))_{n\in\mathbb{Z}}$ is a chain with memory of variable length compatible with the probability tree (τ, p).

Now, by definition, for all $\theta[m,n] \le j \le n$

$$\begin{aligned}
\mathbb{P}(T_{j+1}(\xi_{j+1}) - T_j(\xi_j) > t \mid \xi_{-\infty}^{j}) &= \mathbb{P}\Big(-\frac{1}{q(c_\tau(X_{-\infty}^{j}))}\log \xi_{j+1} > t \;\Big|\; \xi_{-\infty}^{j}\Big) \\
&= \mathbb{P}\Big(\xi_{j+1} < \exp\{-tq(c_\tau(X_{-\infty}^{j}))\} \;\Big|\; \xi_{-\infty}^{j}\Big) \\
&= \exp\{-tq(c_\tau(X_{-\infty}^{j}))\}.
\end{aligned}$$

Therefore, $\mathbb{E}\left[\mathbf{1}\{T_{j+1}(\xi_{j+1}) - T_j(\xi_j) > t\} \mid \xi_{-\infty}^{j}\right]$ is $\mathcal{F}(X(\xi_{-\infty}^{j}))$-measurable and the following equalities are true

$$\begin{aligned}
\mathbb{P}\left(T_{j+1}(\xi_{j+1}) - T_j(\xi_j) > t \mid X(\xi_{-\infty}^{j})\right) &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{T_{j+1}(\xi_{j+1}) - T_j(\xi_j) > t\} \mid \xi_{-\infty}^{j}\right] \mid X(\xi_{-\infty}^{j})\right] \\
&= \mathbb{E}\left[\mathbf{1}\{T_{j+1}(\xi_{j+1}) - T_j(\xi_j) > t\} \mid \xi_{-\infty}^{j}\right] \\
&= \exp\{-tq(c_\tau(X_{-\infty}^{j}))\}.
\end{aligned}$$

□

For the next step we need define in a equivalent way the stopping time $\theta[n]$, for any $n \ge 0$.

Stationarity proof. Define $\xi = (\xi_n)_{n\in\mathbb{Z}} \in [0,1)^{\mathbb{Z}}$. It is enough to show that $(\mathrm{U}(T\xi))_n = (\mathrm{U}(\xi))_{n+1}$ where $T : [0,1)^{\mathbb{Z}} \to [0,1)^{\mathbb{Z}}$ is the translation operator $T(\xi_n) = \xi_{n+1}$. In order to clean the notation, we will denote the translated sequence as $\xi' = (T\xi_j)_{j\in\mathbb{Z}}$ and its respective stopping time as $\theta'[m,n] = \theta[m,n](\xi')$.

Note that, the following sequence of implication are truth

$$\begin{aligned}
\theta[n+1] = m+1 \iff\; & m+1 = \sup\left\{m \le n+1 \;:\; \bigcap_{j=m+1}^{m+N} \{\xi_j \le 1/N\}\right\} \\
\iff\; & m = \sup\left\{m \le n \;:\; \bigcap_{j=m}^{m+N-1} \{\xi'_j \le 1/N\}\right\} \\
\iff\; & \theta'[n] = m.
\end{aligned}$$

Therefore, $\theta'[n] = \theta[n+1] - 1$.

Now, assume that $\mathrm{U}(\xi)_{\theta[n+1]} = \sigma(\{0,\dots,N\})$ and $\mathrm{U}(\xi')_{\theta'[n]} = \sigma'(\{0,\dots,N\})$ where σ and σ' are both permutations of $\{0,1,\dots,N\}$. So, we can suppose, without lost of generality, that, $\mathrm{U}(\xi')_{\theta'[n]} = \sigma\left(\mathrm{U}(\xi)_{\theta[n+1]}\right)$ but then, $\mathrm{U}(\xi')_{\theta[n+1]-1} = \sigma\left(\mathrm{U}(\xi)_{\theta[n+1]}\right)$ for some σ permutation of $\{0,\dots,N\}$.

Suppose that $\mathrm{U}(\xi')_j = \sigma\left(\mathrm{U}(\xi)_{j-1}\right)$, for all $\theta[n+1] \le j \le l$, and remember that $(\mathrm{U}(n))_{n\in\mathbb{Z}}$ is a

Markov chain with memory of length 1, then

$$\begin{aligned}
\mathrm{U}(\xi')_l &= \sum_{j\in\mathcal{N}} \Delta_j(\mathrm{U}(\xi')_{l-1})\mathbf{1}\{\xi'_l \in J^{\mathrm{U}}(j \mid \mathrm{U}(\xi')_{l-1})\} \\
&= \sum_{j\in\mathcal{N}} \Delta_j\left(\sigma(\mathrm{U}(\xi)_l)\right)\mathbf{1}\{\xi_{l+1} \in J^{\mathrm{U}}(j \mid \sigma(\mathrm{U}(\xi)_l))\} \\
&= \sum_{j\in\mathcal{N}} \sigma\left(\Delta_{\sigma^{-1}(j)}(\mathrm{U}(\xi)_l)\right)\mathbf{1}\{\xi_{l+1} \in J^{\mathrm{U}}(\sigma^{-1}(j) \mid \sigma(\mathrm{U}(\xi)_l))\} \\
&= \sigma(\mathrm{U}(\xi)_{l+1}).
\end{aligned}$$

Thus, by the Lemma 2.4, for all $l \geq \theta[n+1]-1$

$$\mathbb{P}\left(\mathrm{U}(\xi')_j \mid \mathrm{U}(\xi')_{j-1}\right) = \mathbb{P}\left(\sigma(\mathrm{U}(\xi)_{j+1}) \mid \sigma(\mathrm{U}(\xi)_j)\right) = \mathbb{P}\left(\mathrm{U}(\xi)_{j+1} \mid \mathrm{U}(\xi)_j\right).$$

Moreover, if $\mathrm{U}(\xi')_l = \sigma(\mathrm{U}(\xi)_{l+1})$ then, $\varphi\big(\sum_{j\in\mathcal{N}} \mathrm{U}(\xi')_l(j)\big) = \varphi\big(\sum_{j\in\mathcal{N}} \mathrm{U}(\xi)_{l+1}(j)\big)$ and $(T_{l+1} - T_l)(\xi') \stackrel{D}{=} (T_{l+2} - T_{l+1})(\xi)$ where $\stackrel{D}{=}$, means the equality in distribution. $\square$

Chapter 3

A model for neural activity in the absence of external stimuli

3.1 Introduction

We study the behavior of a system of interacting neurons in the absence of external stimuli our goal being (i) to determine the long-run behavior of the process with a finite number of neurons, (ii) to study the hydrodynamic limit for this system and (iii) to investigate the possible invariant distributions for the limiting process (the process we obtain by taking the number of neurons to infinity). Our system is composed of N neurons whose state at time $t \geq 0$ is specified by $\mathrm{U}(t) = (\mathrm{U}_1(t), \dots \mathrm{U}_N(t))$, with $\mathrm{U}(t) \in \mathbb{R}_+^N$. For each neuron $i = 1, \dots, N$ and each time $t \geq 0$, $\mathrm{U}_i(t)$ represents the membrane potential of neuron i at time t. We consider two kinds of interactions among neurons and also a constant interplay between neurons and the environment.

More precisely, the neurons interact via *electrical* and *chemical synapses*. Electrical synapses are due to so-called *gap-junction channels* between neurons which induce a constant sharing of potential, pushing the system towards its average value. By contrast, chemical synapses are point events which can be described as follows. Each neuron spikes randomly at rate $\varphi(\mathrm{U}) \geq 0$ which depends on its membrane potential U, we suppose φ a non decreasing function, positive at $\mathrm{U} > 0$ and both integrable and vanishing at 0 (in agreement with the assumption of non external stimuli). When neuron i spikes, its membrane potential is immediately reset to a resting potential 0. Simultaneously, the neurons which are influenced by neuron i receive an additional positive value to their membrane potential. This value may vary for each pair of neurons. Moreover, in the whole time, the neurons loose potential to the environment, due to *leakage channels* which pushes down the membrane potential of each neuron toward zero. This outgoing constant flow of potential is defined as *leak currents.* For technical details we refer the reader to Gersnter and Kistler (2002).

Our system is inspired by the one introduced in Galves and Löcherbach (2013) and De Masi et al. (2014). This model is an example of *piecewise-deterministic Markov processes* (PDPs) introduced by Davis [Dav84]. Such processes combine random jump events, the chemical synapses, with deterministic continuous evolutions, in our case due both to electrical synapses and the leak current. The PDPs have been used also to model neuronal systems by other authors, see for instance the papers [**?**], [DMGLP15], [FL14] and [RT14].

Chemical synapses and leakage make the system non-conservative. Moreover, there is an evident competition between the incoming energy induced by the spikes and the outgoing energy induced by leak currents. Therefore it is natural to ask about the limiting behavior of the system as time $t \to \infty$. The results presented in Theorem 8 and Theorem 9 provide a complete description of the asymptotic distribution of the process with a finite number of neurons. The theorem 8 states that under the presence of the leakage then almost surely there are only a finite number of spikes and the system converges to an "inactive global state" interpreted as "brain sleep". When the leakage is absent we prove that the process is Harris-Recurrent, whenever the initial configuration is non null. This is the content of the Theorem 9.

We then, in the Theorem 10, derive an hydrodynamic equation for the process as the number of neurons N diverges. More specifically, it is shown that the distribution of the membrane potentials becomes deterministic and it is described by a density $\rho_t(u)$ which is proved to obey a nonlinear PDE. For this we work under the mean-field assumption, that is, whenever a neuron spikes it adds the value $1/N$ to the membrane potential of all other neurons.

Finally, we investigate the possible invariant measures for the limiting process. We show, in the Theorem 11, that the limiting process has exactly two invariant measures. In Theorem 12, we provide conditions for extinction or not of this limiting process.

3.2 Model definition and main results

Let $\mathcal{N} = \{1, \cdots, N\}$ be a finite set of neurons, for some fixed integer $N \geq 1$ and consider the family of non-negative *synaptic weights* $(W_{i\to j})_{i,j\in\mathcal{N}} \in \mathbb{R}_+^{\binom{N}{2}}$ such that $W_{i\to i} = 0$ for all $i \in \mathcal{N}$. The value $W_{i\to j}$ corresponds to the value added to the membrane potential of neuron j whenever the neuron i spikes.

We consider a continuous time Markov process

$$\mathrm{U}(t) = (\mathrm{U}_1(t), \ldots, \mathrm{U}_N(t)),\ t \geq 0,$$

taking values in $\mathbb{R}_+^N$, whose infinitesimal generator is given for any smooth test function $f : \mathbb{R}_+^N \to \mathbb{R}$, by

$$\mathcal{L}f(u) = \sum_{i\in\mathcal{N}} \varphi(u_i)[f(\Delta_i(u)) - f(u)] - \lambda \sum_{i\in\mathcal{N}} \left(\frac{\partial f}{\partial u_i}(u)[u_i - \bar{u}] \right) - \alpha \sum_{i\in\mathcal{N}} \left(\frac{\partial f}{\partial u_i}(u)[u_i] \right), \tag{3.1}$$

where, for all $i \in \mathcal{N}$, $\Delta_i : \mathbb{R}_+^N \to \mathbb{R}_+^N$ is defined by

$$(\Delta_i(u))_j = \begin{cases} u_j + W_{i\to j}, & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases},$$

$\lambda, \alpha \geq 0$ are positive parameters modelling, respectively, the strength of electrical synapses and the leakage effect, $\bar{u} = \frac{1}{N}\sum_{i=1}^N u_i$ and

Assumption 3.1. φ is a non-decreasing real-valued function satisfying the conditions $\varphi(0) = 0$ and $\varphi(u) > 0$ for $u > 0$.

The assumption 3.1 implies that external stimuli are not considered. This is a consequence of the condition $\varphi(0) = 0$. In addition, from the neurobiological point of view, it is reasonable to assume that φ is a non-decreasing function since an addition in the membrane potential increases the probability of a spike to occur.

The first term in (3.1) depicts how the chemical synapses are incorporated in our model. Neurons whose potential is u spike at rate $\varphi(u)$. Intuitively this means that for any initial configuration $u \in \mathbb{R}_+^N$ of the membrane potentials

$$\mathbb{P}(\mathrm{U}(t) = \Delta_i(u) \mid \mathrm{U}(0) = u) = \varphi(u_i)t + o(t),\ \text{as } t \to 0$$

Thus, the function φ is called firing or spiking rate of the system.

The second and third terms in (3.1) represent the electrical synapses and the leak current respectively. They describe the deterministic time evolution of the system between two consecutive spikes. More specifically, in an interval of time $[a, b]$, without occurrence of spikes in the whole system, the membrane potential of neuron $i \in \mathcal{N}$ obeys the following ordinary differential equation

$$\frac{d}{dt}\mathrm{U}_i(t) = -\alpha \mathrm{U}_i(t) - \lambda(\mathrm{U}_i(t) - \bar{\mathrm{U}}(t)). \tag{3.2}$$

Notice that the first term of the right-hand side of (3.2) pushes simultaneously all neurons to the resting state, while the second term tends to attract the neurons to the average potential.

Our first theorem proves the existence of the process.

Theorem 6. *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be any function satisfying the Assumption 3.1. For any $N \geq 1$ and any $u \in \mathbb{R}_+^N$ there exists a unique strong Markov process $\mathrm{U}^u(t)$ taking values in $\mathbb{R}_+^N$ starting from u whose generator is given by (3.1).*

Proof. Let $N_i(t),\ t \geq 0$, be the simple point process on R_+ which counts the jump events of neuron $i \in \mathcal{N}$ up to time t. Define $E_i = \sum_{j \neq i} W_{i \to j}$ and $E = \max_{i \in \mathcal{N}} E_i$ and, following De Masi et al. (2014), consider the following random variable, for all $t > 0$,

$$k(t) = \sum_{i \in \mathcal{N}} \int_0^t \mathbf{1}\{\mathrm{U}_i^u(s^-) \leq 2E\} dN_i(s).$$

The random variable $k(t)$ counts the number of spikes of neurons whose the potential is at most $2E$ up to time t.

Suppose U_i fires at time t, in this case

$$\bar{\mathrm{U}}^u(t) = \frac{1}{N} \sum_{j \neq i} \big(\mathrm{U}_j^u(t^-) + W_{i \to j}\big) = \bar{\mathrm{U}}^u(t^-) + \frac{1}{N}\big(E_i - \mathrm{U}_i^u(t^-)\big).$$

Now, using the expression of $\bar{\mathrm{U}}(t)$ above and adapting the proof of Theorem 1 of De Masi et al (2014), we have the following inequalities for all $t > 0$,

$$\bar{\mathrm{U}}^u(t) \leq \bar{\mathrm{U}}^u(0) + \frac{E}{N} k(t), \quad EN(t) \leq N\bar{\mathrm{U}}(0) + 2Ek(t) \quad \text{and} \quad ||\mathrm{U}^u(t)|| \leq (N+1)||\mathrm{U}^u(0)|| + 2Ek(t),$$

where $||\mathrm{U}^u(t)|| = \max_{i \in \mathcal{N}} \mathrm{U}_i^u(t)$.

Since we can bound $Ek(t)$ by a Poisson process of intensity $N\varphi(2E)$, the second inequality above shows that number of jumps of the process is finite almost surely on any finite time interval. To conclude the proof just note that the construction of the process can be achieved by gluing together trajectories given by the deterministic flow between successive jump times. This procedure is feasible since the number of jumps of the process is finite on any finite interval.

□

Now, we shall present an elementary argument which shows that for all leakage rate α large enough and if the firing rate φ is globally Lipschitz with $\varphi(0) = 0$, the system goes extinct. This result was the starting point of this paper. The idea of this proof was taken from discussions with Galves and Löcherbach. The result is the following.

Theorem 7. *For any $N \geq 1$, $\alpha \geq 0$, $\lambda \geq 0$ and c-Lipschitz function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $c > 0$ and $\varphi(0) = 0$, the following inequality holds, for all $t \geq 0$ and $u \in \mathbb{R}_+^N$,*

$$\mathbb{E}[\bar{\mathrm{U}}^u(t)] \leq \bar{u} e^{t(\alpha^* c - \alpha)},$$

where $\alpha^* = \max_{k \in \mathcal{N}} \sum_{j \in \mathcal{N}} W_{j \to k}$. In particular, if $\alpha > \alpha^* c$, then the process goes extinct.

Proof. For each $i \in \mathcal{N}$, plugging $f = \pi_i$ in (3.1), where π_i is the projection onto the i-th coordinate, we have

$$\frac{d}{dt}\mathbb{E}[\mathrm{U}_i^u(t)] = \sum_{j \in \mathcal{N}} W_{j \to i} \mathbb{E}[\varphi(\mathrm{U}_j^u(t))] - \mathbb{E}[\mathrm{U}_i^u(t)\varphi(\mathrm{U}_i^u(t))] - \alpha \mathbb{E}[\mathrm{U}_i^u(t)] - \lambda \mathbb{E}[\mathrm{U}_i^u(t) - \bar{\mathrm{U}}^u(t)].$$

Summing over all $i \in \mathcal{N}$ and then using that φ is a non-negative c-Lipschitz function such that

$\varphi(0) = 0$, it follows that

$$\frac{d}{dt}\mathbb{E}[\bar{\mathrm{U}}^u(t)] \leq (\alpha^* c - \alpha)\mathbb{E}[\bar{\mathrm{U}}^u(t)].$$

Therefore applying the Grownwall's lemma to the inequality above we finish the proof. □

Even assuming that the firing rate φ satisfies only the Assumption 3.1, we claim that for any fixed number of neurons, the presence of the leak current is a necessary and sufficient condition for the extinction of the process. In fact, we shall prove a stronger result. It states that, if there is the leakage, there will be only a finite number of spikes eventually almost surely. On the other hand, it is shown that, excluding the trivial initial configuration, the system is ergodic when there is no leakage. In particular, this results generalize the Theorem 7 above.

In order to state our main result, we need to introduce some extra notation and new assumption.

Assumption 3.2. Assume that there exists a constant $r > 0$ such that

$$\int_0^r \frac{\varphi(u)}{u} du < +\infty.$$

For each neuron $i \in \mathcal{N}$, let $T_1^i = \inf\{s > 0 : \mathrm{U}_i(s) = 0\}$ be the first spiking time of neuron i and for each $k \geq 2$, let $T_k^i = \inf\{s > T_{k-1}^i : \mathrm{U}_i(s) = 0\}$ be the k-th spiking time of neuron i. Then, the first and the k-th spiking time of the system are defined respectively by

$$T_1 = \inf_{i\in\mathcal{N}} T_1^i \quad \text{and} \quad T_k = \inf_{i\in\mathcal{N}, m\geq 1}\{T_k^i > T_{m-1}\}, \quad k \geq 2. \tag{3.3}$$

Our main theorem is given below.

Theorem 8. *Let $(\mathrm{U}^u(t))_{t\geq 0}$ be the Markov process, with $\mathrm{U}^u(0) = u \in \mathbb{R}_+$, whose the infinitesimal generator is given by (3.1) and T_k be as defined in (3.3). Assume that φ satisfies the Assumptions 3.1 and 3.2 with $r > \max_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} W_{j\to i}$. Then for any $\alpha > 0$ and $\lambda \geq 0$,*

$$\mathbb{P}\Big(\sum_{k\geq 1} \mathbf{1}\{T_k < \infty\} < \infty\Big) = 1.$$

Corollary 3.1. *Under the same hypothesis of Theorem 8, for all $i \in \mathcal{N}$ and $u \in \mathbb{R}_+^N$, it holds*

$$\lim_{t\to+\infty} \mathrm{U}_i^u(t) = 0 \quad a.s.$$

In particular, the delta of Dirac at the point 0^N, δ_{0^N}, is the unique invariant measure for the process in the presence of leak currents.

Example 3.1 (Mean-field type of interactions)**.** Suppose that $W_{i\to j} = N^{-1}$ for all $i, j \in \mathcal{N}, i \neq j$. In this case, $\max_i \sum_i W_{i\to j} = (N-1)/N < 1$. In this case the hypothesis of Theorem 8 are fulfilled if, for instance, the spiking rate φ satisfies

$$\int_0^1 \frac{\varphi(u)}{u} du < +\infty.$$

It remains to analyse what happens in the long-run behavior of the system in the absence of the leakage. This is the content of the next result which require an assumption on the interaction graph induced by the synaptic weights $(W_{i\to j})_{i,j\in\mathcal{N}}$. We shall assume also that each neuron influence and it is influenced by at least one other neuron. Formally,

Assumption 3.3. For each neuron $i \in \mathcal{N}$ there exist at least two neurons $j, j' \in \mathcal{N} \setminus \{i\}$ such that $W_{i\to j} > 0$ and $W_{j'\to i} > 0$.

Theorem 9. *Let $(\mathrm{U}^u(t))_{t\geq 0}$ be the Markov process, with $\mathrm{U}^u(0) = u$, whose the infinitesimal generator is given by (3.1) with $\alpha = 0$. Under the Assumptions 3.1 and 3.3, for all $u \in \mathbb{R}_+^N \setminus \{0^N\}$, the process $(\mathrm{U}^u(t))_{t\geq 0}$ is Harris Recurrent. In particular, in this case the process $(\mathrm{U}^u(t))_{t\geq 0}$ does not go extinct.*

3.3 Proof of Theorem 8

In what follows we shall drop de superscript u from $\mathrm{U}^u(t)$ unless some confusion may arises. First of all, observe that, from the equation (3.2), for any time $t \in [T_n, T_{n+1})$,

$$\mathrm{U}_i(t) = \mathrm{U}_i(T_n)e^{-(\alpha+\lambda)(t-T_n)} + \bar{\mathrm{U}}(T_n)e^{-\alpha(t-T_n)}\big(1 - e^{-\lambda(t-T_n)}\big). \tag{3.4}$$

We shall explore this equation many times.

We start giving a lower bound to the probability of having no spikes when the system starts with a initial condition small enough.

Proposition 3.1. *Let $(\mathrm{U}(t))_{t\in\mathbb{R}}$ be the Markov process whose the generator is given by (3.1), $\|\mathrm{U}(t)\| = \max\{\mathrm{U}_i(t), i \in \mathcal{N}\}$ be the maximum potential of the system at time t and T_1 as defined in (3.3). Suppose that φ satisfies the Assumptions 3.1 and 3.2. If $\alpha > 0$, then*

$$\mathbb{P}(T_1 = \infty \mid \|\mathrm{U}(0)\| < r) \geq e^{-\frac{N}{\alpha}\int_0^r \frac{\varphi(u)}{u}du} > 0.$$

Proof. By the equation (3.4), we have, for all $0 \leq t < T_1$, the following inequality

$$\|\mathrm{U}(t)\| \leq r[e^{-(\alpha+\lambda)t} + e^{-\alpha t}(1 - e^{-\lambda t})] = re^{-\alpha t}. \tag{3.5}$$

Using the inequality above and the non-decreasing assumption on φ, we have

$$\begin{aligned}\mathbb{P}(T_1 > t \mid \|\mathrm{U}(0)\| < r) &\geq \exp\left\{-\int_0^t N\varphi(\|\mathrm{U}(s)\|)ds\right\} \\ &\geq \exp\left\{-\frac{N}{\alpha}\int_{re^{-\alpha t}}^r \frac{\varphi(u)}{u}du\right\}. \end{aligned}\tag{3.6}$$

Therefore, taking t to infinite, the result follows. □

Lemma 3.1. *For any fixed $u \in \mathbb{R}_+^N$, with $\|u\| > R_0 > 0$, there exists $t_0 = t_0(u) > 0$ such that $\forall t \geq t_0$ it holds that*

$$\mathbb{P}\big(\|\mathrm{U}(t)\| \leq R_0 \mid \mathrm{U}(0) = u\big) \geq \exp\{-N\varphi(\|u\|)t_0\} > 0 \tag{3.7}$$

Proof. First of all, observe that by equation (3.4), for all $i \in \mathcal{N}$, we get for $t < T_1$

$$\|\mathrm{U}_i(t)\| \leq \big(\|u_i\| - \bar{u}\big)e^{-\alpha t} + \bar{u}e^{-\alpha t} = \|u_i\|e^{-\alpha t}.$$

Taking $t_0 = -\frac{1}{\alpha}\log\frac{R_0}{\|u\|}$ we have that, $\forall t \geq t_0$

$$\mathbb{P}\big(\|\mathrm{U}(t_0)\| \leq R_0 \mid \mathrm{U}(0) = u\big) \geq \mathbb{P}\big(T_1 > t_0 \mid \mathrm{U}(0) = u\big) \geq \exp\{-N\varphi(\|u\|)t_0\}.$$

□

Proposition 3.2. *Under the same hypothesis of Proposition 3.1. For $\theta < (\alpha+\lambda)^{-1}$, it holds that*

$$\mathbb{P}(T_1 > \theta \mid \|\mathrm{U}(0)\| > r) \leq e^{-\theta\varphi(\beta)} < 1,$$

where $\beta = \beta(\theta) = r(1-(\alpha+\lambda)\theta) > 0$.

Proof. Indeed,

$$\mathbb{P}(T_1 > \theta \mid \|\mathrm{U}(0)\| > r) \quad \leq \quad \exp\left\{-\int_0^\theta \varphi(\|\mathrm{U}(s)\|)ds\right\}. \tag{3.8}$$

Moreover, by (3.4) for all $0 < s \leq \theta$ is true that

$$\|\mathrm{U}(s)\| \geq \|\mathrm{U}(0)\|e^{-(\alpha+\lambda)s} \geq re^{-(\alpha+\lambda)s} \geq r(1-(\alpha+\lambda)\theta). \tag{3.9}$$

From (3.8) and (3.9) we deduce the desired inequality

$$\mathbb{P}(T_1 > \theta \mid \|\mathrm{U}(0)\| \geq r) \leq e^{-\theta\varphi(\beta)} < 1.$$

□

Remark 3.1. To prove the last result we use only the lower bound of $\|\mathrm{U}(0)\|$. So that the same result remain true if we suppose a upper bound for it. Thereafter, when necessary we shall use the last proposition in the form

$$\mathbb{P}\big(T_1 > \theta \mid r < \|\mathrm{U}(0)\| \leq R_0\big) \leq e^{-\theta\varphi(\beta)} < 1.$$

The next result claims that even when the maximum potential is large, but smaller than R_0, there is a positive probability of all potentials become smaller that r after a fix time $T > 0$.

Before state the result, we shall define some new variables. Let $(S_m)_{m\geq 1}$ be the sequence of spiking neurons of the system. The event $\{S_m = i\}$ means that the $\mathrm{U}_i(T_m) = 0$. Besides, for a given u, define the sequence $v_k = v_k(u)$, in the following way: $v_0 = u$ and for $k = 1, \ldots, N$,

$$(v_k)_i = \Big(u_i + \sum_{j=1}^{k} W_{j\to i}\Big)\mathbf{1}_{\{i>k\}} + \sum_{j=i+1}^{k} W_{j\to i}\mathbf{1}_{\{i<k\}} \; i = 1\ldots, N.$$

Remark 3.2. An important fact here is that, by definition, $\|v_N(u)\| \leq \max_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} W_{j\to i}$, for any $u \in \mathbb{R}_+^N$.

We shall prove the next lemma for a specific sequence of spikes, but the same result remain true for any sequence $S_1, \ldots, S_N$ such that $\{S_1, \ldots, S_N\} = \mathcal{N}$. The proof can be adapted changing the sequence $v(u)$ depending on the order of the spikes.

Lemma 3.2. *For any fixed $u \in \mathbb{R}_+^N$ conditioning to the event $S = \{S_k = k,\, k = 1, \ldots, N\}$*

$$\mathrm{U}_i^u(T_k) = \big(v_k(u)\big)_i + R_\theta(u, T_1^k),\; k = 1\ldots, N,$$

where $T_1^k = (T_1 \ldots, T_k)$ and $R_\theta(u, T_1^k)$ is a function of $T_1^k = (T_1, \ldots T_k)$ and u which goes to zero when $\theta \to 0$.

Proof. The proof is made by induction in k and since u is fixed, we shall omit the argument u in R_θ and v. Define, for commodity $T_0 = 0$, and consider the event

$$M_\theta = \Big\{T_k - T_{k-1} < \theta,\, k = 1, \ldots, N\Big\}.$$

In words, M_θ is the event in which the first k inter-spikes intervals have length smaller than θ.

Note that, given $M_\theta \cap S$, making use of the series expansion of the exponential function in the equation (3.4), we have that $\mathrm{U}_1(T_1) = 0$ and we can write, for all $i = 2, \dots, N$,

$$\begin{aligned} \mathrm{U}_i(T_1) &= W_{1\to i} + \big(u_i - \bar{u}\big)\big(1 + R_\theta(T_1)\big) + \bar{u}\big(1 + R_\theta(T_1)\big) \\ &= (v_0)_i + R_\theta(T_1) \end{aligned}$$

Suppose that, for a fixed $k = 2, \dots, N-1$, the lemma is true. So, in particular we have that $\bar{\mathrm{U}}(T_k) = \bar{v_k} + R_\theta(T_1^k)$. Now, by hypothesis, $\mathrm{U}_{k+1}(T_{k+1}) = 0$ and, using the same arguments as before, for all $i \neq k$

$$\begin{aligned} &\mathrm{U}_i(T_{k+1}) \\ &\quad = W_{k+1\to i} + \big((v_k)_i - \bar{v_k} + R_\theta(T_1^k)\big)\big(1 + R_\theta(T_1^{k+1})\big) + (\bar{v_k} + R_\theta(T_1^k))\big(1 + R_\theta(T_1^{k+1})\big) \\ &\quad = (v_{k+1})_i + R_\theta(T_1^{k+1}) \end{aligned}$$

□

Proposition 3.3. *Consider* $(\mathrm{U}(t))_{t\in\mathbb{R}}$ *the Markov process whose the generator is given by (3.1),* $\|\mathrm{U}(t)\|$ *and* T_1 *as in proposition 3.1. Suppose that* φ *satisfies the Assumption 3.1 and* $\alpha > 0$*. If* $r > \max_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} W_{j\to i}$*, then there exists* $T > 0$ *such that, for all* $R_0 > 0$*,*

$$\mathbb{P}\big(\|\mathrm{U}(T)\| < r \mid r \leq \|\mathrm{U}(0)\| \leq R_0\big) > 0.$$

Proof. Define $\tau_0 = 0$ and for $k \geq 1$, τ_k is the k-th spike of the neuron with the largest potential, formally,

$$\tau_k = \inf\Big\{t > \tau_{k-1} : \bigcup_{i\in\mathcal{N}} \{\|\mathrm{U}(t^-)\| = \mathrm{U}_i(t^-), \mathrm{U}_i(t) = 0\}\Big\}. \tag{3.10}$$

Now, take $\theta < (\alpha + \lambda)^{-1}$ (this is possible since $\alpha > 0$). By the proposition 3.2, and its Remark, it holds that

$$\mathbb{P}\big(T_1 < \theta, T_1 = \tau_1 \mid r \leq \|\mathrm{U}(0)\| \leq R_0\big) \geq \Big(1 - e^{-\theta\varphi(\beta)}\Big)\frac{1}{N}.$$

Define $R_\theta = \max\{R_\theta(u, T_1^k) : \|u\| \leq R_0\}$ and, if necessary, take a smaller θ in such a way that $r > \max_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} W_{j\to i} + R_\theta$ (it is possible since $R_\theta(u, T_1^k) \to 0$ when $\theta \to 0$). Now, consider $T_0 = 0$, $T = N\theta$ and the following sequence of events

$$A_1 = \{T_1 < \theta, T_1 = \tau_1, \|\mathrm{U}(T_1)\| < r, T_2 > T\},$$

and for $l = 2 \dots, N$,

$$A_l = \Big\{\cap_{k=1}^{l} \{T_k < T_{k-1} + \theta, T_k = \tau_k, \}, \cap_{k=1}^{l-1}\{\|\mathrm{U}(T_k)\| > r\}, \|\mathrm{U}(T_l)\| < r, T_{l+1} > T\Big\}.$$

The event A_k corresponds to the following situation. At the first $k-1$ consecutive spikes of the maximum, there always exists at least one neuron whose potential is larger than r. But at k-th spike of the maximum of all membrane potentials get less than r and no more spikes happen up to time T.

From the definitions of θ and the events A_k we have

$$\{\|\mathrm{U}(T)\| < r\} \supseteq \bigcup_{k=1}^{N} A_k.$$

Wherefore, it suffices to compute the probability of each event A_k above, conditioned to the event $\{\|\mathrm{U}(0)\| > r\}$. Now, defining $\beta_0 = \big(1 - e^{-\theta\varphi(\beta)}\big)\frac{1}{N}$, and $\beta_1 = \exp\Big\{-\frac{N}{\alpha}\int_{re^{-\alpha N\theta}}^{r} \frac{\varphi(u)}{u} du\Big\}$, by

the proposition 3.2 and the inequality (3.6), it follows that

$$\mathbb{P}(A_1 \mid r \leq \|\mathrm{U}(0)\| \leq R_0) > \beta_0\beta_1\mathbb{P}(\|\mathrm{U}(T_1)\| < r \mid r \leq \|\mathrm{U}(0)\| \leq R_0, T_1 < \theta, T_1 = \tau_1).$$

Similarly, we have

$$\begin{aligned}\mathbb{P}(A_2 \mid r \leq \|\mathrm{U}(0)\| \leq R_0) > \beta_0^2\beta_1\mathbb{P}(\|\mathrm{U}(T_1)\| > r \mid \|r \leq \|\mathrm{U}(0)\| \leq R_0, T_1 < \theta, T_1 = \tau_1)\\ \times\mathbb{P}(\|\mathrm{U}(T_2)\| < r \mid \|\mathrm{U}(T_1)\| > r, T_2 < T_1 + \theta, T_2 = \tau_2).\end{aligned}$$

Thus, summing the two inequalities above and using that $A_1 \cap A_2 = \emptyset$ and that $\mathbb{P}(A) + \mathbb{P}(A^c)\mathbb{P}(B) \geq \mathbb{P}(B)$, we get the following lower bound for $A_1 \cup A_2$,

$$\mathbb{P}(A_1 \cup A_2 \mid r \leq \|\mathrm{U}(0)\| \leq R_0) > \beta_0^2\beta_1\mathbb{P}(\|\mathrm{U}(T_2)\| < r \mid \|\mathrm{U}(T_1)\| > r, T_2 < T_1 + \theta, T_2 = \tau_2).$$

Proceeding in this way for the other terms, we obtain that

$$\begin{aligned}&\mathbb{P}(\|\mathrm{U}(T)\| < r \mid r \leq \|\mathrm{U}(0)\| \leq R_0)\\ &\quad > \beta_0^N\beta_1\mathbb{P}\left(\|\mathrm{U}(T_N)\| < r \mid \|\mathrm{U}(T_{N-1})\| > r, \cap_{k=1}^N\{T_k < T_{k-1} + \theta, T_k = \tau_k\}\right).\end{aligned}$$

On the other hand, by definition of A_N, is clear that $A_N \subseteq M_\theta \cap S$, where M_θ and S were defined in the proof of Lemma 3.2. Then by the Lemma 3.2 and the choose of θ, it follows that

$$\mathrm{U}_i(T) \leq \sum_{j=i+1}^{N} W_{j\to i} + R_\theta < r, \ i = 1, \ldots, N-1, \text{ and } \mathrm{U}_N(T) = R_\theta < r$$

so that,

$$\mathbb{P}\left(\|\mathrm{U}(T)\| < r \mid \|\mathrm{U}(T_{N-1})\| > r, \cap_{k=1}^N\{T_k < T_{k-1} + \theta, T_k = \tau_k\}\right) = 1.$$

□

As a consequence of proposition 3.3 we have

Corollary 3.2. *Le $T > 0$ be the positive constant given by the proposition 3.3. Define the stopping times $R_1 = \inf\{n \geq 1; \|\mathrm{U}(nT)\| \leq r\}$ and $R_k = \inf\{n \geq R_{n-1}; \|\mathrm{U}(nT)\| \leq r\}$ for all $k \geq 2$. Then, under the same hypothesis of the proposition 3.3,*

$$\sum_{k=1}^{\infty} \mathbf{1}\{R_k < \infty\} = \infty \quad \mathbb{P} - a.s.$$

Proof of Theorem 8. For this proof we must define the following stopping times: $K_1 = \inf\{n \geq 1 : T_n > R_1T\}$, $J_1 = \inf\{n \geq 1 : R_nT > T_{S_1}\}$, and for $k \geq 2$, $K_k = \inf\{n > J_{k-1} : T_n > R_{J_{k-1}}T\}$ and $J_k = \inf\{n \geq 1 : R_nT > T_{S_{k-1}}\}$. From corollary 3.2 all these stopping times are well defined. Now, from Theorem 6 and the definition of K_k

$$1 = \mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{i=1}^{\infty}\{T_i > R_kT\} \,\middle|\, \bigcap_{i=1}^{\infty}\{T_i < \infty\}\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty}\{K_k < \infty\} \,\middle|\, \bigcap_{i=1}^{\infty}\{T_i < \infty\}\right).$$

Thus, $\cap_{k=1}^{\infty}\{T_k < \infty\}$ implies $\cap_{k=1}^{\infty}\{K_k < \infty\}$.

On the other hand, by proposition 3.1, one knows that

$$\mathbb{P}\left(\bigcap_{k=1}^{n}\{K_k < \infty\}\right) \leq \left(1 - e^{-\frac{N}{\alpha}\int_0^r \frac{\varphi(u)}{u}du}\right)^n,$$

which converges to 0 when n diverges. □

3.4 Proof of Theorem 9

To simplify the proof of the theorem 3.4 we shall split it into several steps. The main part of the argument is to find a *recurrent regeneration set* B in sense that

(i) if we write $T_B = \inf\{t > 0 : \mathrm{U}^u(t) \in B\}$, then $\mathbb{P}(T_B < \infty | \mathrm{U}(0) = u) = 1$, for each $u \in \mathbb{R}_+^N \setminus \{0^N\}$, and

(ii) there exist $t_* > 0$, $\epsilon > 0$ and a probability measure ν on $\mathbb{R}_+^N$ such that

$$P_{t_*}(u, A) := \mathbb{P}(\mathrm{U}^u(t_*) \in A) \geq \epsilon\nu(A), \ u \in B,$$

for all measurable set $A \in \mathcal{B}(\mathbb{R}_+^N)$.

Usually, Markov processes with a regeneration set are called *Harris chains.* For such processes an invariant measure always exits, see for instance [Asm03].

In what follows, for any positive real number $a > 0$ we use the notation $R_a(x)$, meaning that there exists a constant $l > 0$ such that $|R_a(x)| \leq la$ for all x. When a function satisfies such condition it is called a function of order a. Note that we are not specifying the domain in which the function R is defined on.

For each $\epsilon > 0$ define the following event

$$M_\epsilon = \{T_1 < \epsilon, T_k - T_{k-1} < \epsilon, \ k = 2, \cdots, N\},$$

and consider again the event $S = \{S_k = k, \, k = 1, \ldots, N\}$.

The first lemma below says that, conditioning on the event $M_\epsilon \cap S$, when ϵ is sufficiently small the process evolves, modulo an error of small order, as in the case without electrical synapses ($\lambda = 0$). Before stating this lemma we need to introduce a finite sequence of potential configurations.

Consider the sequence $(v_k)_{k=0,\ldots,N}$ with $v_k \in R_+^N$ given by

$$v_0 = \Big(\sum_{j=2}^N W_{j\to 1}, \sum_{j=3}^N W_{j\to 2}, \ldots, W_{N\to N-1}, 0\Big), \tag{3.11}$$

and for $1 \leq k \leq N$, $(v_k)_k = 0$ and for $i \neq k$,

$$(v_k)_i = \Big(\sum_{j=i+1}^k W_{j\to i}\Big)\mathbf{1}_{\{i<k\}} + \Big((v_0)_i + \sum_{j=1}^k W_{j\to i}\Big)\mathbf{1}_{\{i>k\}}.$$

Lemma 3.3. *Fix $\delta > 0$. If $\mathrm{U}(0) = u \in B(v_0, \delta)$, then conditioning on $M_\epsilon \cap S$, for each $k = 1, \cdots, N$, the following equalities hold:*

(i) $\mathrm{U}_i(T_k) = (v_k)_i + \sum_{r=i+1}^k \lambda(T_r - T_{r-1})d_i(r-1) + R_{\delta\epsilon}(T_1^k, u) + R_{\epsilon^2}(T_1^k, u)$, *if $i < k$;*

(ii) $\mathrm{U}_i(T_k) = (v_k)_i + \sum_{r=1}^k \lambda(T_r - T_{r-1})d_i(r-1) + R_\delta(u) + R_{\delta\epsilon}(T_1^k, u) + R_{\epsilon^2}(T_1^k, u)$, *if $i > k$;*

(iii) $\bar{\mathrm{U}}(T_k) = \bar{v}_k + R_\delta(u) + R_\epsilon(T_1^k, u)$, *if $k < N$ and* $\bar{\mathrm{U}}(T_N) = \bar{v}_0 + R_\epsilon(T_1^N, u)$,

where $d_i(m) = \bar{v}_m - (v_m)_i$, $T_0 = 0$ and $T_1^k = (T_1, \ldots, T_k)$. Furthermore, all the partial derivatives of the remainder functions $R_{\delta\epsilon}(T_1^k, u)$, $R_{\epsilon^2}(T_1^k, u)$ above are either of order δ or ϵ.

Proof. The proof is given by induction on k. On the event $M_\epsilon \cap S$, we have that $\mathrm{U}_1(T_1) = 0$ and for each $i = 2, \ldots, N$ and $\mathrm{U}(0) = u \in B(v_0, \delta)$,

$$\begin{aligned} \mathrm{U}_i^u(T_1) &= W_{1\to i} + \bar{\mathrm{U}}(0) + (1 - \lambda T_1 + R_{\epsilon^2}(T_1))\big((v_0)_i + R_\delta(u) - \bar{\mathrm{U}}(0)\big) \\ &= (v_1)_i + \lambda T_1 d_i(0) + R_\delta(u) + R_{\epsilon\delta}(T_1, u) + R_{\epsilon^2}(T_1, u), \end{aligned}$$

where in the first equality we have used the expansion series of the exponential function.

Thus, from the expression of $\mathrm{U}_i^u(T_1)$ above, we conclude that

$$\begin{aligned}\bar{\mathrm{U}}(T_1) &= \bar{v}_1 + \lambda T_1 \frac{1}{N}\sum_{i=2}^{N} d_i(0) + R_\delta(u) + R_{\epsilon\delta}(T_1,u) + R_{\epsilon^2}(T_1,u)\\ &= \bar{v}_1 + R_\delta(u) + R_\epsilon(T_1,u).\end{aligned}$$

In addition, it is easy to check that all remainders functions above have partial derivatives with respect to T_1 and all of them are either R_δ or R_ϵ functions. Therefore (i), (ii) and (iii) it is verified for $k=1$.

Now, suppose that (i), (ii) and (iii) hold for some fixed $1<k<N$. As before, on the event $M_\epsilon \cap S$, we have $\mathrm{U}_{k+1}^u(T_{k+1})=0$ and, for $i<k+1$, by the inductive hypothesis,

$$\begin{aligned}\mathrm{U}_i^u(T_{k+1}) &= W_{k+1\to i} + \bar{\mathrm{U}}(T_k) + (1-\lambda(T_{k+1}-T_k) + R_{\epsilon^2}(T_{k-1},T_k))(\mathrm{U}_i(T_k)-\bar{\mathrm{U}}(T_k))\\ &= (v_{k+1})_i + \sum_{r=i+1}^{k+1}\lambda(T_r-T_{r-1})d_i(r-1) + R_{\delta\epsilon}(T_1^{k+1},u) + R_{\epsilon^2}(T_1^{k+1},u)\end{aligned}$$

Using the same argument for the case when $i>k+1$, we get *(ii)* for $k+1$. Using again the inductive hypothesis and looking at the expression written in the first equality of the membrane potential, it is readily seen that the remainder functions possesses partial derivatives with to T_l, for $l=1,\dots,k+1$ and they are either of order δ or ϵ.

Finally, summing $\mathrm{U}_i(T_{k+1})$ over all neurons $i=1,\dots,N$

$$\bar{\mathrm{U}}(T_{k+1}) = \bar{v}_{k+1} + R_\epsilon(T_1^{k+1},u) + R_\delta(u).$$

□

Note that $v_N = v_0$, thus, from the previous lemma it follows the

Corollary 3.3. *Under the same assumptions of Lemma 3.3, if $T=N\epsilon<T_{N+1}$ then for each $i=1,\dots,N$, the following equality is verified*

$$\mathrm{U}_i^u(T) = (v_0)_i + \lambda(T-T_N)d_i(0) + \sum_{r=i+1}^{N}\lambda(T_r-T_{r-1})d_i(r-1) + R_{\delta\epsilon}(T_1^N,u) + R_{\epsilon^2}(T_1^N,u).$$

Remark 3.3. In order to simplify the notation, we shall denote the map $\gamma^0 : M_\epsilon \to \mathbb{R}_+^N$ by $\gamma^0(t_1^N) = (\gamma_1^0(t_1^N),\dots,\gamma_N^0(t_1^N))$ where, $\gamma_N^0(t_1^N) = (v_0)_N + \lambda(T-t_N)d_N(0)$ and for each $i=1,\dots,N-1$,

$$\gamma_i^0(t_1^N) = (v_0)_i + \lambda(T-t_N)d_i(0) + \sum_{r=i+1}^{N}\lambda(t_r-t_{r-1})d_i(r-1).$$

By Corollary 3.3, conditioning on the event $M_\epsilon \cap S \cap \{T<T_{N+1}\}$, $T=N\epsilon$, we have the following representation for all $\mathrm{U}(0)=u\in B(v_0,\delta)$,

$$\mathrm{U}^u(T) = \gamma^0(T_1^N) + R_{\delta\epsilon}(T_1^N,u) + R_{\epsilon^2}(T_1^N,u),$$

where both $R_{\delta\epsilon}(T_1^N,u)$ and $R_{\epsilon^2}(T_1^N,u)$ are multivalued functions whose the L_1-norms are remainders functions of order $\delta\epsilon$ and ϵ^2 respectively.

Define, for all $k\in\mathcal{N}$, $m_k = \sum_{j\in\mathcal{N}} W_{k\to j}$ the total amount of potential the neuron k insert in the system each time it has a spike. We need this notation a cup of time in the proofs of the this chapter. Moreover, by the Assumption 3.3 we know that $m_k>0$ for any $k\in\mathcal{N}$. Using this and the fact that $\bar{v}_k \geq m_k/N$ we get the following corollary

Corollary 3.4. *For each $u \in B(v_0, \delta)$, the absolute value of the determinant of the Jacobian of map $M_\epsilon \ni t_1^N \mapsto \gamma_u(t_1^N) = \gamma^0(t_1^N) + R_{\delta\epsilon}(t_1^N, u) + R_{\epsilon^2}(t_1^N, u)$, is given by*

$$|J\gamma_u(t_1^N)| = \lambda^N \prod_{i=1}^N \bar{v}_i + R_\epsilon(t_1^N, u) + R_\delta(t_1^N, u),$$

which, under the Assumption 3.3, is different from zero for δ and ϵ small enough for all $(t_1^N) \in M_\epsilon$ and $u \in B(v_0, \delta)$.

We shall use this representation to show that our process, at time $T = N\epsilon$, satisfies a localized Doeblin condition (see proposition 3.4). Before proving this proposition, we need an extra lemma. Here again the non-decreasing assumption on φ is important.

Lemma 3.4. *Let $f_u(t_1, \dots, t_N) = f_{u,(T_1,\dots,T_N),(S_1=1,\dots,S_N=N)}(t_1, \dots, t_N)$ denote the joint density of $(T_1, \dots, T_N)$ with $(S_1, \dots, S_N)$ restricted to the event S, when the starting configuration is u. Under the Assumption 3.3, it holds that for any $0 < \delta < (v_0)_1$ there exists a constant $C_1 > 0$ such that for all $u \in B(v_0, \delta)$,*

$$f_u(t_1, \dots, t_N) \geq C_1, \text{ for } (t_1, \dots, t_N) \in M_\epsilon.$$

In particular, there exist $\epsilon > 0$ and a constant $C_2 > 0$ such that for $u \in B(v_0, \delta)$ and $T = N\epsilon$,

$$P_T(u, B(v_0, \delta)) \geq C_2 \epsilon^N > 0.$$

Proof. (i) Since $\mathbb{P}_u(T_1 > t) = \exp\left[-\int_0^t \sum_{j=1}^N \varphi(\mathrm{U}_j^u(s))ds\right]$, we immediately see that the density function of T_1 with $S_1 = 1$ given that $\mathrm{U}(0) = u$ is

$$f_{T_1,S_1=1|u}(t_1) = \varphi(\mathrm{U}_1^u(t_1)) \exp\left[-\int_0^{t_1} \sum_{j=1}^N \varphi(\mathrm{U}_j^u(s))ds\right], \text{ for } t_1 \geq 0. \tag{3.12}$$

Since $u \in B(v_0, \delta)$ and $\delta < (v_0)_1$ (we can find such δ because the Assumption 3.3 implies $(v_0)_1 > 0$), we know that there exist positive constants c_1^1 and c_2^1 such that $c_1^1 < u_1$ and for all $j = 1, \dots, N$, $u_j < c_2^1$, thus $c_1^1/N < \bar{u} < c_2^1$. But then for all $u \in B(v_0, \delta)$ and $j = 1, \dots, N$, we have that $\mathrm{U}_j(s) = \bar{u}(1 - e^{-\lambda s}) + u_j e^{-\lambda s} < c_2^1$, for all $0 \leq s < T_1$. Thus, using that φ is non-decreasing, from the previous inequality and the identity (3.12) it follows that $f_{T_1,S_1=1|u}(t_1) \geq \varphi(c_1^1)e^{-t_1 N\varphi(c_2^1)}$.

Now, from the definition of the process one easily sees that the density function of the increment T_2 with $S_2 = 2$ given $T_1 = t_1, S_1 = 1$ and $\mathrm{U}(0) = u$, for $t_2 > t_1$, is

$$f_{T_2,S_2=2|T_1=t_1,S_1=1,u}(t_2) = \varphi(\mathrm{U}_2^{\Delta_1(\mathrm{U}^u(t_1))}(t_2)) \exp\left[-\int_{t_1}^{t_2} \sum_{j=1}^N \varphi(\mathrm{U}_j^{\Delta_1(\mathrm{U}^u(t_1))}(s))ds\right]. \tag{3.13}$$

Note that if we denote $K = \max_{i,j\in\mathcal{N}} W_{i\to j}$ then, for all $j \in \mathcal{N}$,

$$\Delta_1(\mathrm{U}^u(t_1))_j = W_{1\to j} + \mathrm{U}_j^u(t_1^-) < K + c_2^1 := c_2^2.$$

Moreover, we have that $\bar{\mathrm{U}}(T_1) > m_1/N$ which is greater than zero by the Assumption 3.3. Therefore, for all $t_1 < t_2 < T_2$,

$$\mathrm{U}_2(t_2) > (m_1/N)(1 - e^{-\lambda(t_2-t_1)}) := c_1^2,$$

From these two inequalities, using again the monotonicity of φ and that the average potential is constant between successive jumps it follows that there exist positive constants c_1^2 and c_2^2 such that $f_{T_2,S_2=2|T_1=t_1,S_1=1,u}(t_2) \geq \varphi(c_1^2)e^{-(t_2-t_1)N\varphi(c_2^2)}$, for $t_2 > t_1$. Proceeding in this manner we obtain

sequences $(c_1^n)_{n=1,\dots,N-1}$ and $(c_2^n)_{n=1,\dots,N-1}$ satisfying for $k = 2, \dots, N$,

$$f_{T_k,S_k=k|T_{k-1}=t_{k-1},S_{k-1}=k-1,\dots,T_1=t_1,S_1=1,u}(t_k) \geq \varphi(c_1^{k-1})e^{-(t_k-t_{k-1})N\varphi(c_2^{k-1})},$$

where $t_k > t_{k-1} > \dots > t_1 \geq 0$. Thus, we have that, over M_ϵ, the product of these conditional densities is strictly positive, finishing the proof. □

Proposition 3.4. *(Localized Doeblin Condition) Under the Assumption 3.3, for any $u \in B(v_0, \delta)$, there exists a non negative function h_u such that for all measurable sets $A \in \mathcal{B}(\mathbb{R}_N^+ \setminus \{0^N\})$,*

$$P_T(u, A) \geq \int_A h_u(v)dv.$$

Moreover, there exist a measurable set I, with positive Lebesgue measure, and a constant $C_3 > 0$ such that $h_u(v) \geq C_3 1_I(v)$ for all $u \in B(v_0, \delta)$.

Proof. For each $u \in B(v_0, \delta)$, as in Corollary 3.4 let us call $\gamma_u : M_\epsilon \to I_u$, the map

$$\gamma_u(t_1^N) = \gamma^0(t_1^N) + R_{\delta\epsilon}(t_1^N) + R_{\epsilon^2}(t_1^N),$$

where $I_u = \gamma_u(M_\epsilon)$. From the corollaries 3.3, 3.4 and the remark 3.3, it follows that for each $u \in B(v_0, \delta)$, conditioning to $M_\epsilon \cap S$, the random vector $\mathrm{U}^u(T)$ has a density h_u, where

$$h_u(v) = \begin{cases} f_u(g_u(v)) \mid Jg_u(v) \mid & , \text{ if } v \in I_u \\ 0 & \text{otherwise,} \end{cases}$$

with $g_u : I_u \to M_\epsilon$ being the inverse of γ_u. This concludes the first part of the proposition.

The proof of second part is more delicate and requires some work. From Corollary 3.4 and Lemma 3.4 part (i) it suffices to prove that there exists a set I such that $I \subset \cap_{u\in B(v_0,\delta)} I_u$.

Now, consider the event B_ϵ defined by

$$B_\epsilon = S \cap \Big\{(i-1)\epsilon + \frac{\epsilon}{4} < T_i < i\epsilon - \frac{\epsilon}{4},\ i = 1, \cdots, N\Big\},$$

define $I = \gamma_0(B_\epsilon)$ and fix $v \in I$. We want to show that for all $u \in B(v_0, \delta)$ there is an vector $t_1^N = t_1^N(u) = (t_1(u), \dots, t_N(u)) \in M_\epsilon$ such that $\gamma_u(t_1^N) = v$. To this end, we introduce the function $F(s, t_1^N) = v - \gamma_0(t_1^N) - s[R_{\delta\epsilon}(t_1^N, u) + R_{\epsilon^2}(t_1^N, u)]$, for $s \in [0, 1]$ and $t_1^N \in M_\epsilon$. Note that we need to show the existence of vector $t_1^N \in M_\epsilon$ such that $F(1, t_1^N) = 0$. This means that we need to study the equation $F(s, t_1^N) = 0$ with t_1^N as function of s. To ease the notation, from now on we will write t instead of t_1^N.

Note that by the definition of I, there exists $t_0 \in B_\epsilon$ such that $F(0, t_0) = 0$. Besides, $t = t(s)$ is solution of the equation $F(s, t) = 0$ if and only if it satisfies

$$0 = -D\gamma_0(t(s)) \cdot \frac{dt(s)}{ds} - [R_{\delta\epsilon}(t(s), u) + R_{\epsilon^2}(t(s), u)] - s[DR_{\delta\epsilon}(t(s), u) + DR_{\epsilon^2}(t(s), u)] \cdot \frac{dt(s)}{ds}$$

or equivalently,

$$\Big[- D\gamma_0(t(s)) - s\big(DR_{\delta\epsilon}(t(s), u) + DR_{\epsilon^2}(t(s), u)\big)\Big] \cdot \frac{dt(s)}{ds} = R_{\delta\epsilon}(t(s), u) + R_{\epsilon^2}(t(s), u), \tag{3.14}$$

where $Df(\cdot)$ stands for the differential operator of f. By corollary 3.4, the linear operator inside the brackets is invertible and the function on the right hand side is of order $\delta\epsilon + \epsilon^2$ whose the derivative is of order $\delta + \epsilon$. Therefore, it follows that $t = t(s)$ is a solution of (3.14) if and only if it is the solution of the ODE of the form

$$\frac{d}{ds}t(s) = H(s, t(s)),\ t(0) = t_0, \tag{3.15}$$

where the derivative of H with respect to t exists and it is of order $\delta+\epsilon$. In particular, it is limited and therefore the ODE has unique solution $t=t(s)$ for all s such that $t(s)\in M_\epsilon$. Since $t(s)$ moves as H which is of order $\delta\epsilon+\epsilon^2$ and the initial condition t_0 is at a distance of order ϵ of M_ϵ, we have that $t(1)\in M_\epsilon$, decreasing both δ and ϵ if necessary. Hence, we have that $I\subset\cap_{u\in B}I_u$. □

Now, in order to obtain the desired regenerative set B it remains to control excursion out of $B=B(v_0,\delta)$. We start proving

Lemma 3.5. *There exists a measurable set A such that $\sup_{u\in A}\mathbb{E}[T_A^+\mid \mathrm{U}(0)=u]<\infty$, where $T_A^+=\{t>T_{A^c}:\mathrm{U}(t)\in A\}$ is the time of the first return to A after an exit, and*

$$\mathbb{E}[T_A\mid \mathrm{U}(0)=u]\le C(u_1+\ldots+u_N)<\infty,$$

for all $u\in A^c$, where C is a positive constant. Moreover, $B=B(v_0,\delta)\subseteq A$ for all δ sufficiently small and there exists a constant $D>0$ such that $A\subseteq[0,D]^N$.

Proof. Define $V(u)=\sum_{i=1}^N u_i$, for all $u\in\mathbb{R}_+^N$, and $m_i=\sum_{j\neq i}W_{i\to j}$. We notice then that $\mathcal{L}V(u)=\sum_{i=1}^N\varphi(u_i)[m_i-u_i]$, where $\mathcal{L}$ is the generator of the process $(\mathrm{U}(t))_{t\ge0}$ given in (3.1). Set $\mathcal{L}V(v_0)=L$, recall the definition of v_0 in (3.11), and put $a=\min\{-1,2L\}$. We claim that the set $A=\{u\in\mathbb{R}_+^N:\mathcal{L}V(u)>a\}$ fulfils the desired conditions.

Indeed, since $M_t=V(\mathrm{U}(t))-V(\mathrm{U}(0))-\int_0^t\mathcal{L}V(\mathrm{U}(s))ds$ is a martingale with $M_0=0$, and T_A is a stopping time, we have that $M_{t\wedge T_A}$ is a martingale and

$$\mathbb{E}[V\big(\mathrm{U}(t\wedge T_A)\big)\mid \mathrm{U}(0)=u]-V(u)=\mathbb{E}\Big[\int_0^{t\wedge T_A}\mathcal{L}V(\mathrm{U}(s))ds\mid \mathrm{U}(0)=u\Big]. \tag{3.16}$$

Now, take $u\in A^c$, so that $\mathbb{E}\Big[\int_0^{t\wedge T_A}\mathcal{L}V(\mathrm{U}(s))ds\mid \mathrm{U}(0)=u\Big]\le a\mathbb{E}[t\wedge T_A\mid \mathrm{U}(0)=u]$ which, together with (3.16), implies that $\mathbb{E}[t\wedge T_A\mid \mathrm{U}(0)=u]\le CV(u)$, for all $t\ge0$, where $C=-a^{-1}>0$. Since $\mathbb{E}[t\wedge T_A\mid \mathrm{U}(0)=u]\ge t\mathbb{P}(T_A>t\mid \mathrm{U}(0)=u)$, it follows that $\mathbb{P}(T_A<\infty\mid \mathrm{U}(0)=u)=1$ and that $\mathbb{E}[T_A\mid \mathrm{U}(0)=u]\le CV(u)=C(u_1+\ldots+u_N)$.

Consider the configuration $u=(m_1/2,\ldots,m_N/2)$, then $\mathcal{L}V(u)=\sum_{i\in\mathcal{N}}\varphi(m_i/2)m_i/2>0$. Since $\mathcal{L}V$ is a continuous function and $\lim_{u\to\infty}\mathcal{L}V(u)=-\infty$, it follows that there exists $D>0$ such that $A\subseteq[0,D]^N$. Since $\mathcal{L}V(v_0)=L>a$, using again the continuity of $\mathcal{L}V$ we conclude that $B(v_0,\delta)\subset A$ for all δ small enough.

Finally, given $u\in A$, call $v=\mathrm{U}^u(T_{A^c})$. Since $||u||\le D$, we have that $||v||\le D+\max_{i,j}W_{i\to j}:=K$, so that by the strong Markov property

$$\mathbb{E}[T_A^+\mid \mathrm{U}(0)=u]\le\sup_{v\notin A:|v|\le K}\mathbb{E}[T_A\mid \mathrm{U}(0)=v]\le CNK.$$

□

We shall now use the lemma above to prove

Lemma 3.6. *Let $B=B(v_0,\delta)$. For all δ sufficiently small and any $u\in\mathbb{R}_+^N\setminus\{0^N\}$,*

$$\mathbb{P}[T_B<\infty\mid \mathrm{U}(0)=u]=1. \tag{3.17}$$

The proof is made using the same arguments as in the Lemma 3.4 and for this reason we shall be more direct.

Proof. Notice that at the first spike time T_1,

$$||\mathrm{U}(T_1)||\ge\bar{\mathrm{U}}(T_1)=\frac{1}{N}\sum_{i=1}^N\mathrm{U}_i(T_1)\ge\frac{1}{N}\min\{m_i,i\in\mathcal{N}\}:=b>0.$$

Since for each $u \in \mathbb{R}_+^N \setminus \{0^N\}$, $\mathbb{P}(T_1 < \infty \mid \mathrm{U}(0) = u) = 1$, by the Markov property we may prove (3.17) for all $u \in \mathbb{R}_+^N$ such that $u_i > b$, for some $i = 1, \dots, N$. Moreover, since the process $(\mathrm{U}(t))_{t\geq 0}$ reaches A almost surely (the set A given by Lemma 3.16) we may check the equality (3.17) for all $u \in A$ such that $u_i > b$, for some $i = 1, \dots, N$. Therefore, we will prove that (3.17) holds for all $u \in A$ with $u_1 > b$, being the other cases treated likewise.

Define $T_0 = 0$ and consider the event $B_\epsilon = \{T_{k-1} + \epsilon/2 < T_k < T_{k-1} + \epsilon, S_k = k, k = 1 \dots, N\}$.

Since $u \in A$ with $u_1 > b$, for all $s < T_1$, $\mathrm{U}_1(s) > b/N$ and, by the Lemma 3.5, $\mathrm{U}_j^u(s) < D$, for any $j \in \mathcal{N}$. Now, observe that, as in Lemma 3.4, the joint density of T_1 and $S_1 = 1$ given that $\mathrm{U}(0) = u \in A$, is the function described in 3.12. Therefore, if t_1 satisfies the conditions on B_ϵ, then $f_{T_1,S_1=1|u}(t_1) > \varphi(b/N)e^{-(\epsilon/2)N\varphi(D)} > 0$.

If $T_1 = t_1 > \epsilon/2$ we have that $\mathrm{U}_2(t_1) > b/n(1 - e^{-t(\epsilon/2)}) + W_{1\to 2} := c_1^2 > 0$, $\bar{\mathrm{U}}(t_1) > m_1/N$ and, for all $j \in \mathcal{N}$, it holds that $\mathrm{U}_j(s) < D + K := c_2^2$. Arguing as in the Lemma 3.4, we consider the joint density of T_2 and $S_2 = 2$ given that $T_1 = t_1, S_1 = 1$ and $\mathrm{U}(0) = u \in A$, defined in 3.13 can be bounded below by $\varphi(c_1^2)e^{-(\epsilon/2)\varphi(c_2^2)}$ if t_1 and t_2 fulfils the conditions asked in B_ϵ.

Applying the argument other $N-2$ times we show that there exist two sequences $(c_1^n)_{n=1,\dots,N-1}$ and $(c_2^n)_{n=1,\dots,N-1}$ such that, for $n = 2, \dots, N$, if $(t_1, \dots, t_k)$ fulfils the conditions imposed in B_ϵ

$$f_{T_k,S_k=k|T_{k-1}=t_{k-1},S_{k-1}=k-1,\dots,T_1=t_1,S_1=1,u}(t_k) > \varphi(c_1^{k-1})e^{-(\epsilon/2)N\varphi(c_2^{k-1})}.$$

Therefore, under B_ϵ, the product of these conditional densities is strictly positive and, for any $u \in A$, it holds that

$$P_T\big(u, B(v_0, \delta)\big) > 0,$$

which finishes the proof. □

Proof of Theorem 3.4. It is an immediate consequence of Proposition 3.4 and Lemma 3.6.

□

3.5 Hydrodynamic Limit under Mean-Field Assumption

We now study what happens with the interacting system $\mathrm{U}^N(t)$ as $N \to \infty$. For simplicity, we will assume that all neurons behave similarly, leading to a mean field description. This means we assume that for all $i \neq j$, $W_{i\to j} = N^{-1}$. We stress that the model considered here with $\alpha = 0$ (that is without leakage currents) is exactly the one studied in De Masi et al. (2014). Therefore, it is straightforward to apply the techniques developed in that paper to derive a limit equation for the system as $N \to 0$. The limit equation is obtained by showing that the distribution of membrane potentials becomes deterministic and it is described by a limit density function $\rho_t(u)$ which is proved to obey the non linear PDE

$$\frac{\partial \rho_t(u)}{\partial t} + \frac{\partial [V(u, \rho_t)\rho_t(u)]}{\partial u} = -\varphi(u)\rho_t(u), \quad t > 0, u > 0, \tag{3.18}$$

where $V(u, \rho_t) = -\alpha u - \lambda(u - \bar{u}_t) + p_t$, and for each $t \geq 0$,

$$\bar{u}_t = \int_0^\infty u\rho_t(u)du, \qquad p_t = \int_0^\infty \varphi(u)\rho_t(u)du \tag{3.19}$$

are respectively the limit average potential and the limit spiking rate of the system. The boundary conditions of (3.18) are specified by

$$\rho_0(u) = v_0(u), \ \rho_t(0) = v_1(t), \tag{3.20}$$

$v_0(u)$ is determined by the problem $v_0(u) = \psi_0(u)$, while $v_1(t)$ has to be derived together with (3.18). It turns out that the expression of $v_1(t)$ is given by

$$v_1(t) = \frac{p_t}{\lambda \bar{u}_t + p_t}, \tag{3.21}$$

so that the function $\rho_t(u)$ may not be continuous, unless that $\psi(0) = \frac{p_0}{\lambda \bar{u}_0 + p_0}$. Thus a weak formulation of (3.18) is needed.

Definition 4. A real valued function $\mathbb{R}_+ \times \mathbb{R}_+ \ni (t,u) \mapsto \rho_t(u)$ is a weak solution of (3.18)-(3.20) if for all smooth functions $\phi(u)$, $\mathbb{R}_+ \ni t \mapsto \int_0^\infty \phi(u)\rho_t(u)du$ is continuous in t, differentiable in $t > 0$ and

$$\begin{aligned} &\frac{\partial}{\partial t}\int_0^\infty \phi(u)\rho_t(u)du - \int_0^\infty \phi'(u)V(u,\rho_t)\rho_t(u)du - \phi(0)V(0,\rho_t)v_1(t) \\ &= -\int_0^\infty \varphi(u)\phi(u)\rho_t(u)du, \\ &\int_0^\infty \phi(u)\rho_0(u)du = \int_0^\infty \phi(u)v_0(u)du, \end{aligned} \tag{3.22}$$

where $V(u,\rho_t) = -u\alpha - \lambda(u - \bar{u}_t) + p_t$, with $\bar{u}_t$ and p_t as in (3.19).

The solution of (3.22) can be computed explicitly by the method of characteristics. Characteristics are curves along which the PDE reduces to an ODE. They are defined by the equation

$$\frac{dx(t)}{dt} = V(x(t),\rho_t).$$

The solution of the ODE above on the interval $[s,t]$, with value u at s is denoted by $T_{s,t}(u)$, $u \in \mathbb{R}_+$, and it has the following expression:

$$T_{s,t}(u) = e^{-(\alpha+\lambda)(t-s)}u + \int_s^t e^{-(\alpha+\lambda)(t-h)}[\lambda \bar{u}_h + p_h]dh. \tag{3.23}$$

Before stating precisely the result, we recall that the process is denoted by

$$\mathrm{U}^N(t) = (\mathrm{U}_1^N(t), \ldots, \mathrm{U}_N^N(t)), t \geq 0,$$

and notice also that we may identify $\mathrm{U}^N(t)$ with its empirical distribution

$$\mu_t^N = N^{-1}\sum_{i=1}^N \delta_{\mathrm{U}_i^N(t)}. \tag{3.24}$$

In this way, the process $\mathrm{U}^N(t), t \geq 0$, may be viewed as an element of a Skorokhod space $t \mapsto \mu_t^N \in D(\mathbb{R}_+, \mathcal{S}')$, where $\mathcal{S}$ is the Schwartz space of all smooth functions $\phi : \mathbb{R}_+ \to \mathbb{R}$. For any fixed $T > 0$, we denote the restriction of the process to $[0,T]$ by $\mu_{[0,T]}^N \in D([0,T],\mathcal{S}')$ and write $\mathcal{P}_{[0,T]}^N$ to denote its law on $D([0,T],\mathcal{S}')$. Finally we assume that

Assumption 3.4. For all N, $\mathrm{U}_i^N(0)$ are i.i.d random variables distributed according to $\psi_0(u)du$ on $\mathbb{R}_+$. The density function ψ_0 has compact support $[0,R_0]$ and also satisfies

(i) $\psi_0 > 0$ on $[0,R_0)$ and $\psi_0 \equiv 0$ on $[R_0,\infty)$.

(ii) $\psi(u) \geq c(u - R_0)^2$, $c > 0$ in an left neighbourhood of R_0.

The result is the following.

Theorem 10. *Grant Assumptions 3.1 and 3.4, for any fixed $T > 0$,*

$$\mathcal{P}^N_{[0,T]} \xrightarrow{w} \mathcal{P}_{[0,T]}, \tag{3.25}$$

(weak convergence in $D([0,T],\mathcal{S}')$) as $N \to \infty$, where $\mathcal{P}_{[0,T]}$ is the law on $D([0,T],\mathcal{S}')$ supported by the distribution valued trajectory ω_t given by

$$\omega_t(\phi) = \int_0^\infty \phi(u)\rho_t(u)du, \quad t \in [0,T],\ \phi \in \mathcal{S}.$$

Here, $\rho_t(u)$ is the unique weak solution of (3.18)-(3.20) with $v_0 = \psi_0$ and v_1 given by (3.21). Moreover, $\rho_t(u)$ is a continuous function of (t,u) in $\mathbb{R}_+ \times \mathbb{R}_+ \setminus \left\{(T_{0,t}(0),t) : t \in R_+\right\}$ where it is differentiable in t and u and its derivatives satisfy (3.18). Furthermore, for any $t \geq 0$, $\rho_t(u)$ has compact support in u and

$$\rho_t(0,r) = \frac{p_t}{\lambda_r \bar{u}_t + p_t} \text{ and } \int_0^\infty \rho_t(u)du = 1.$$

Its explicit expression for $u \geq T_{0,t}(0)$, is:

$$\rho_t(u) = \psi_0\left(T^{-1}_{0,t}(u)\right)\exp\left\{-\int_0^t \left[\varphi\left(T^{-1}_{s,t}(u)\right) - \alpha - \lambda\right]ds\right\},$$

and for $u = T_{s,t}(0)$ for some $0 < s \leq t$,

$$\rho_t(u) = \frac{p_s}{p_s + \lambda\bar{u}_s}\exp\left\{-\int_s^t [\varphi\left(T_{s,h}(0)\right) - \alpha - \lambda]dh\right\}.$$

Proof. The proof is, with minor modifications, analogous to the one provided in the Theorem 2 of [DMGLP15]. □

In the next section, we investigate the possible invariant for the limiting process.

3.6 Long-run behavior of the Limiting Process

In this section we study the long-run behavior of the limiting process. The solution $\rho_t(u)$ of (3.18)-(3.20) can be interpreted as the density of the membrane potential U_t of a single neuron evolving within a system of infinite neurons. The process $(\mathrm{U}(t))_{t\geq 0}$ is a time-inhomogeneous Markov process whose time dependent generator $L = (L_t)_{t\geq 0}$ is such that for any $t \geq 0$ and any smooth function $f : \mathbb{R}_+ \to \mathbb{R}$,

$$L_t f(u) = \varphi(u)[f(0) - f(u)] + f'(u)[p_t - \alpha u - \lambda(u - \bar{u}_t)]. \tag{3.26}$$

We start investigating the possible invariant measures for the limiting process $(\mathrm{U}(t))_{t\geq 0}$. It is shown that $(\mathrm{U}(t))_{t\geq 0}$ possesses exactly two invariant measures.

Theorem 11. *Grant Assumption (3.1), let $\alpha, \lambda \geq 0$. The process $(\mathrm{U}(t))_{t\geq 0}$ admits exactly two invariant measures. The first one is δ_0. The second one is given by*

(i) if either $\alpha > 0$ or $\lambda > 0$, we have that

$$\rho(u) = \frac{p}{p - \lambda(u - \bar{u}) - \alpha u}\exp\left(-\int_0^u \frac{\varphi(v)}{p - \lambda(v - \bar{u}) - \alpha v}dv\right)\mathbf{1}_{\left\{0 \leq u < \frac{p+\lambda\bar{u}}{\alpha+\lambda}\right\}}, \tag{3.27}$$

where $p > 0$ and $\bar{u} > 0$ are such that $\int_0^\infty \varphi(u)\rho(u)du = p$ and $\int_0^\infty u\rho(u)du = \bar{u}$.

(ii) If $\alpha = \lambda = 0$, then

$$\rho(u) = \exp\left(-\frac{1}{p}\int_0^u \varphi(v)dv \right), \tag{3.28}$$

where $p > 0$ is such that $\int_0^\infty \varphi(u)\rho(u)du = p$.

In either cases, $\int_0^\infty \rho(u)du = 1$.

Proof. It follows from Theorem 8 of [FL14]. □

By Theorem 11 we have that the limiting process admits two invariant regimes and in what follows we investigate the behavior of the system $(\mathrm{U}(t))_{t\geq 0}$ as $t \to \infty$.

Assumption 3.5. (i) $\varphi \in C^2(\mathbb{R}_+)$ is a convex increasing function satisfying

$$\sup_{u\geq 1}[\varphi'(u)/\varphi(u) + \varphi(u)/\varphi'(u)] < \infty.$$

(ii) It holds that $\limsup_{u\to\infty}[\varphi'(u)/\varphi(u) < 1]$.

(iii) There are $\xi \geq 1, \zeta \geq \xi - 1$ and some constants $0 < c < C$ such that

$$cu^\xi \leq \varphi(u) \leq C(u^{\xi-1} + u^\zeta).$$

Theorem 12. *Grant Assumption (3.1), let $\alpha, \lambda \geq 0$ and $(\mathrm{U}(t))_{t\geq 0}$ be the the process whose generator is given by (3.26).*

(i) Assume additionally Assumption 3.5-(i) and that the function ψ_0 appearing in Assumption 3.4 satisfies $\psi_0(0) = 1$, $\int_0^\infty \varphi^2(u)\psi_0(u)du < \infty$ and $\int_0^\infty |\psi_0'(u)|du < \infty$. Let $\alpha = \lambda = 0$ and denote by $\rho(t)$ the law of $\mathrm{U}(t)$ and write $\rho(du) = \rho(u)du$ for the invariant probability measure defined in (3.28). Then we have

$$\lim_{t\to\infty} ||\rho(t) - \rho||_{TV} = 0,$$

where $||\cdot||_{TV}$ denotes the total variation distance. In particular, the process does not go extinct almost surely.

(ii) Let $\alpha = 0$ and $\lambda > 0$ and assume Assumptions 3.5 itens (i), (ii) and (iii). Suppose also that $\int_0^\infty u\psi_0(u)du > 0$ and $\int_0^\infty u^{\zeta+1}\psi_0(u)du < \infty$. Then $\mathrm{U}(t)$ does not go to 0 as $t \to \infty$.

(iii) (Stability of the invariant measure δ_0) Let $\alpha > 0$ and $\lambda \geq 0$ and let R_α be a positive number such that $\varphi(u) \leq u\alpha/2$, for all $u \leq R_\alpha$. If $\varphi'(0) < \alpha$ and the support of ψ_0 is included in $[0, R_\alpha]$, then $\mathrm{U}(t)$ goes to 0 exponentially fast. In particular, $\mathrm{U}(t)$ goes extinct.

(iv) Let $\alpha > 0$ and $\lambda \geq 0$. If $\varphi'(0) > \alpha$ and $\int_0^\infty u\psi_0(u)du > 0$, then $\mathrm{U}(t)$ does not go to 0 as $t \to \infty$.

Proof. (i)-(ii) It follows respectively from the Proposition 9 and Proposition 11 of [FL14].

(iii) Indeed, since $\mathrm{U}(t) = 0$ at each spiking time $T_n = t$ and between consecutive spikes $\mathrm{U}(t)$ follows (3.23), then

$$\mathrm{U}(t) \leq e^{-(\alpha+\lambda)t}\mathrm{U}(0) + \int_0^t e^{-(\alpha+\lambda)(t-h)}[\lambda\bar{u}_h + p_h]dh.$$

Now let r_t be the the rightmost point of the support of $\rho_t(u)$. Then, from inequality above we deduce that

$$r_t \leq e^{-(\alpha+\lambda)t}r_0 + \int_0^t e^{-(\alpha+\lambda)(t-h)}[\lambda r_h + \varphi(r_h)]dh. \tag{3.29}$$

Define $\tau = \inf\{t > 0 : r_t > R_\alpha\}$. For $t < \tau$, we have from (3.29) that

$$r_t \leq e^{-(\alpha+\lambda)t} r_0 + \int_0^t (\lambda + \alpha/2) e^{-(\alpha+\lambda)(t-h)} r_h dh.$$

By iterating the above inequality n times we have that there exists a constant $C > 0$ such that

$$r_t \leq e^{-(\alpha+\lambda)t} r_0 \Big(\sum_{k=0}^{n-1} \frac{[(\alpha/2+\lambda)t]^k}{k!} \Big) + C \frac{[(\alpha/2+\lambda)t]^n}{n!},$$

so that by taking $n \to \infty$ we have $r_t \leq e^{-\alpha/2t} r_0$. Since $r_0 \leq R_\delta$, it follows that $\tau = \infty$ and $\log(\mathrm{U}(t)) \leq \log(r_t) \leq -\alpha/2$, for all $t \geq 0$, concluding the proof of (iii).

(iv) Since $\varphi'(0) > \alpha$, we may find $\delta > 0$ and $R_\delta > 0$ small enough such that for all $u \leq R_\delta$, $\varphi(u) - \alpha u > \delta u$. Decreasing R_δ if necessary, we may suppose that for all $u \leq R_\delta$,

$$\varphi(u)(1-u) - \alpha u > c\delta u, \tag{3.30}$$

where c is a positive constant sufficiently small.

Now, suppose by contradiction that $\mathrm{U}(t)$ goes in law to 0 as $t \to \infty$. In particular, we have that $E[\mathrm{U}(t)] \to 0$ as $t \to \infty$. As a consequence of the continuity of $\rho_t(u)$ we may deduce that here exists t_0 such that for all $t \geq t_0$, $r_t \leq R_\delta$, where r_t is the rightmost point of the support of $\rho_t(u)$.

Finally, notice that

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[\mathrm{U}(t)] &= \mathbb{E}[L(\mathrm{U}(t))] \\ &= \mathbb{E}[\varphi(\mathrm{U}(t))] - \mathbb{E}[\varphi(\mathrm{U}(t))\mathrm{U}(t)] - \alpha\mathbb{E}[\mathrm{U}(t)] \\ &= \int_0^\infty [\varphi(u)(1-u) - \alpha u]\rho_t(u)du. \end{aligned} \tag{3.31}$$

Thus, from (3.30) and (3.31), we get that $\frac{d}{dt}\mathbb{E}[\mathrm{U}(t)] \geq c\delta\mathbb{E}[\mathrm{U}(t)]$, and therefore, $\mathbb{E}[\mathrm{U}(t)] \geq e^{c\delta t}\mathbb{E}[\mathrm{U}(0)]$. Therefore, since $\mathbb{E}[\mathrm{U}(0)] > 0$, it follows $\mathrm{U}(t)$ can not go in law to 0 as $t \to \infty$, which is a contradiction.

□

Chapter 4

Hidden Variable Length Markov Chains

Is the brain a statistician? The Predictive coding principle postulates that neural networks learn the statistical regularities inherent in the natural world and reduce redundancy by removing the predictable components of the input, transmitting only what is not predictable [HR11]. The neuroscience community has provided considerable evidences suggesting that the brain at different scales is consistent with the predictive coding principle. For instance, at the microscopic scale, neurons in the visual cortex respond to different types of stimuli depending on their spatial location. Neurons in the primary visual cortex (V1) respond to bars of different orientation, while those located at areas V2 and V4 respond to more complex shapes. This selectiveness effect can be understood by means of a hierarchical model of predictive coding [RB99]. Indeed, knowing that visual system is hierarchically structured and that its connections are reciprocal, Rao and Ballard proposed that higher-level areas predict lower-level input through feedback connections, whereas lower-level areas signal the difference between actual neural activity and the higher-level predictions [JRBB06]. After training this neural network model on image patches taken from natural scenes, they have found that the response selectivities of the model neurons resembled the ones of the neurons in the visual system, corroborating the predictive coding hypothesis.

At the macroscopic scale the predictive coding has been investigated, for instance, in the rhythmic perception domain by means of electroencephalographic (EEG) recording of brain signals [WLvW+11] [WCD12]. The signal collected through this technique is a summation of electrical activity in the brain and contains spontaneous fluctuations that complicate the analysis of a specific effect of interest (Buzski, 2006). Traditionally, experimental protocols try to overcome this difficulty with a large number of repetitions of the same stimulus (Lopes da Silva, 2006). The average of the attempts associated with that stimulus presentation allows to identify a specific response (event related response, ERP). Using this approach, it was shown that auditory stimuli are associated with evoked potentials occurring between 50ms and 200ms after their presentation that predominate in the primary auditory regions (electrodes T3, T4, T7, T8). Also, the expectation of the occurrence of an auditory stimulus generates a modulation of brain activity happening around 200 ms after the activity commonly evoked by the presence of the stimulus. This activity, recorded predominantly in frontocentral electrodes in the scalp, was called "mismatch negativity" (Ntnen et al., 2007; Bendixen, 2009). The mismatch negativity is frequently interpreted in terms of predictive coding. The auditory system would acquire an internal model of regularities in the auditory inputs that are used to generate weighted predictions about the incoming stimuli. If these predictions differ from actual stimulus, it results in a mismatch signal [WCD12].

All the results presented in this chapter are motivated by the investigation of the predictive coding at the rhythmic perceptual level within the framework of stochastic process. Specifically, it is proposed a stochastic modelling-based paradigm to address the question of how to retrieve structural features of stochastic rhythmic sources from EEG data. This is done through a case study in which the EEG data is driven by two different random sources whose rhythmic random

sequences are produced by first considering a fixed structure - (2,1,1) or (2,1,0,1) - the symbol 2 meaning a strong beat, 1 a weak beat and 0 a silence unit, and then omitting each weak beat 1 independently of the past with a fixed probability. We stress that the mathematical framework developed here may be applied to many others situations.

To formulate mathematically the question posed above, we present a suitable class of stochastic processes, herein called *hidden variable length Markov chains* (HVLMC), to model EEG signals evoked by random rhythmic structures. Next, we propose a consistent statistical procedure to perform statistical model selection in the class of hidden variable length Markov chains. This is the main contribution of this chapter. We also present a simulation study where we apply this statistical procedure in a example.

Informally, a hidden variable length Markov chain is a bivariate stochastic process $\big((X_n, Y_n)\big)_{n\in\mathbb{Z}}$ in which:

(i) the state sequence $(X_n)_{n\in\mathbb{Z}}$, usually not observed, is a stochastic chain with memory of variable length compatible with a probabilistic context tree (τ, p) and

(ii) the observable sequence $(Y_n)_{n\in\mathbb{Z}}$ is such that the distribution of Y_{n+1} given the joint past history $(X_{-\infty}^n, Y_{-\infty}^n) = (x_{-\infty}^n, y_{-\infty}^n)$ depends only on $c_\tau(x_{-\infty}^n)$, the context in τ associated to the sequence of symbols $x_{-\infty}^n$.

Notice that the item (ii) above implies, in particular that, conditionally on the state sequence $(X_n)_{n\in\mathbb{Z}}$, the observable process $(Y_n)_{n\in\mathbb{Z}}$ is a sequence of independent variables.

To the best of our knowledge, the hidden variable length Markov models appeared first in Wang and Liu in 2006 as a novel and general approach for time-series data mining [WZF+06]. As an application in real data, Wang and Liu applied this approach to mine four kinds of patterns from 3D motion capture data, which is typical for the high-dimensionality and complex dynamics. A slightly different version of HVLMC's have been recently studied in [Dum14]. In Dumont's version, the state sequence $(X_n)_{n\in\mathbb{Z}}$ is a stochastic chain with memory of variable length (as above) while the observable sequence $(Y_n)_{n\in\mathbb{Z}}$ is such that the distribution of Y_{n+1} given the joint past history $(X_{-\infty}^n, Y_{-\infty}^n) = (x_{-\infty}^n, y_{-\infty}^n)$ depends only on x_n, rather than $c_\tau(x_{-\infty}^n)$. Under these assumptions, Dumont has proposed an estimator of the context tree of the hidden Markov process, which needs no prior upper bound on the depth of the context tree, proving its consistency trough information-theoretic mixtures inequalities.

The notion of hidden variable length Markov models provided above naturally generalize the hidden Markov models introduced by Baum and co-authors in the late 1960s. Indeed, the hidden Markov models are obtained in the particular case in which the state sequence $(X_n)_{n\geq 0}$ is a k-order Markov chain for some positive integer number k. Herein instead, the state sequence is considered to be a variable length Markov chain, a more parsimonious stochastic process introduced in 1983 by Rissanen. For a didactic introduction on hidden Markov models with applications in speech recognition we refer the reader to Rabiner (1889).

In our case study, the state process $(X_n)_{n\in\mathbb{Z}}$ plays the role of the rhythmic sources, while the observable process $(Y_n)_{n\in\mathbb{Z}}$ plays the role of the associated EEG signals driven by the auditory stimuli produced by these sources. In this scenario, the item (ii) means that the EEG activity driven by two stimuli coming from the same auditory sound, but occupying distinct positions in the structure, may have different laws. In other terms, we postulate that the EEG signal is not a function of the acoustic features of the auditory stimuli, it depends actually on the position relative to the structure occupied by these auditory stimuli.

Our problem of finding structural characteristics in the EEG signals means we have to deal with a "context tree estimation" in functional data. Our statistical procedure borrows the ideas both from the Algorithm Context introduced by Rissanen in 1983 and random projections introduced in [CAFR06] by Cuesta-Albertos and co-authors. In a nutshell, the method of random projections states that, under suitable conditions, a distribution on an infinite dimensional space is determined by just one randomly chosen projection. Instead of facing the problem related to infinite-dimensional objects, we choose at random a Gaussian process, project the objects in it,

solve the problem, and, then, translate the solution to the original infinite-dimensional space. Doing this we replace our sample of EEG data to an one dimensional sample. The way the Algorithm Context works can be summarized as follows. Given a sample produced by a chain with variable memory, we start with a maximal tree of candidate contexts for the sample. The branches of this first tree are then pruned starting from the leaves towards the root until we obtain a minimal tree of contexts well adapted to the sample [GL08]. Our method consists in applying the algorithm context to the one dimensional projected data obtained after projecting the EEG data in a randomly selected Gaussian direction.

4.1 Random rhythmic sources and Hidden VLMC Definition

We shall now formulate precisely the question of how to retrieve structural features of a random source through EEG signals driven by a sample produced by this source. We start describing how to generate a sample of each one of the *rhythmic random sources* considered here. Both sources produce random sequences of *strong beats*, *weak beats* and *silence units*. We encode by 2 a strong beat, 1 a weak beat and 0 a silence unit, so that the alphabet of two sources is the set $A = \{0, 1, 2\}$. The first stochastic rhythmic source is called *Ternary* or *Waltz*. Its algorithm of the generation can be describe as follows:

(i) Consider the deterministic sequence $(w_n)_{n\in\mathbb{Z}}$

$$\ldots 2\ 1\ 1\ 2\ 1\ 1\ 2\ 1\ 1\ 2 \ldots$$

(ii) Fix $\epsilon > 0$ and denote by $(\xi_n)_{n\in\mathbb{Z}}$ a sequence of i.i.d Bernoulli random variables such that

$$\mathbb{P}(\xi_n = 0) = 1 - \mathbb{P}(\xi_n = 1) = \epsilon. \tag{4.1}$$

The parameter ϵ is the omission probability of the weak beats. The ternary random source $(W_n)_{n\in\mathbb{Z}}$ is then defined by

$$W_n = G(w_n, \xi_n),$$

where the function $G : A \times \{0, 1\} \to A$ is such that $G(2, \xi_n) = 2$ and $G(1, \xi_n) = \xi_n$. This means that the Ternary random source $(W_n)_{n\in Z}$ is obtained by first considering a deterministic sequence $(w_n)_{n\in\mathbb{Z}}$ and then replacing in a i.i.d way, with probability ϵ, each weak beat 1 by a silence unit 0. Observe, in particular, that all strong beats are never erased. It is easily checked that the process $(W_n)_{n\in Z}$ defined above is a stochastic chain with memory of variable length in the sense of Definition 1. Its context tree is depicted in figure 4.1(a), and its family of transition probabilities is given in Table 4.1 below. For instance, it is easy to check that the transition probabilities

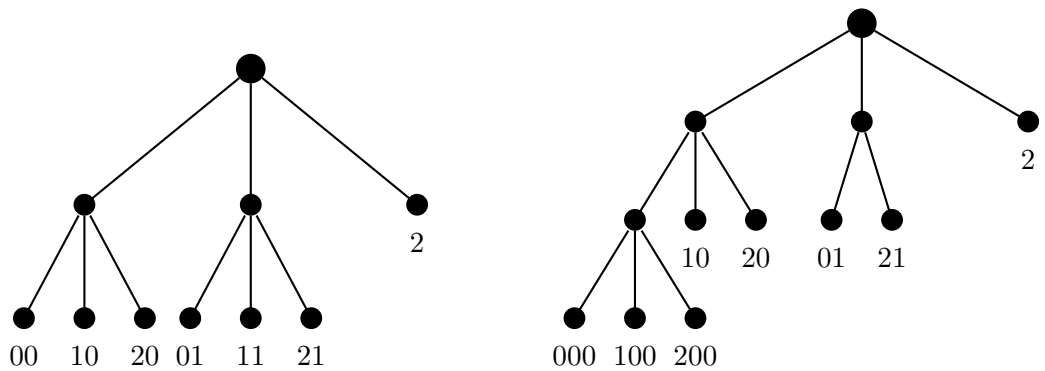
Contexts(w)	**p(0|w)**	**p(1|w)**	**p(2|w)**
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
11	0	0	1
01	0	0	1
20	ϵ	$1-\epsilon$	0
10	0	0	1
00	0	0	1

Table 4.1: *Contexts and transition probabilities on the alphabet $A = \{0, 1, 2\}$, for the ternary random source.*

associated to context 2 are $p(0|2) = \epsilon,\ p(1|2) = 1 - \epsilon$ and $p(2|2) = 0$. Indeed, in the deterministic sequence $(w_n)_{n\in\mathbb{Z}}$ after each a strong beat 2 there is always a weak beat 1 which can be erased, becoming a 0, with probability ϵ or kept with probability $1 - \epsilon$. It is not difficult also to compute the transition probabilities associated to contexts 00 and 10:

$$p(0|00) = p(0|01) = 0,\ p(1|00) = p(1|01) = 0 \text{ and } p(2|00) = p(2|01) = 1.$$

In fact, both strings 00 and 01 were originally a string 11 which had, consecutively, the two symbols 1 and only the second symbol 1 erased by chance. Moreover, in the deterministic sequence after two consecutive weak beats 11 there is always a strong beat 2 which is never erased.



(a) Ternary context tree (b) Quaternary context tree

Figure 4.1: *Context trees of the rhythmic random sources*

The second stochastic rhythmic source is called *Quaternary* or *Samba*. This source can be generated through the following algorithm:

(i) Consider the deterministic sequence $(s_n)_{n\in\mathbb{Z}}$

$$\ldots 2\ 1\ 0\ 1\ 2\ 1\ 0\ 1\ 2\ 1\ 0\ 1\ 2\ldots$$

(ii) Let the parameter $\epsilon > 0$ and the process $(\xi_n)_{n\in\mathbb{Z}}$ defined as in (4.1). The Quaternary random source $(S_n)_{n\in Z}$ is then defined by

$$S_n = H(s_n, \xi_n),$$

where the function $H : A\times\{0,1\} \to A$ is such that $H(2,\xi_n) = 2$, $H(0,\xi_n) = 0$ and $H(1,\xi_n) = \xi_n$. In other terms, the Quaternary random source $(S_n)_{n\in\mathbb{Z}}$ is obtained by first considering a deterministic sequence $(s_n)_{n\in\mathbb{Z}}$ and then replacing in a i.i.d way each weak beat 1 for a silence unit 0, with probability ϵ.

As in the Ternary source, all strong beats 2 are not erased. However, differently from the ternary source, the symbol 0 may be either a *constitutive silence* or a *erased weak beat*. The constitutive silence corresponds to the case $H(0,\xi_n) = 0$, while erased weak beats correspond to the cases $H(1,\xi_n) = \xi_n$ in which $\xi_n = 0$. In other words, the symbol 0 in the Samba source may have been the symbol 0 or a symbol 1 which was erased, in the deterministic sequence. Moreover, it may be easily checked that $(S_n)_{n\in\mathbb{Z}}$ defined above is a stochastic chain with memory of variable length whose the context tree is depicted in the Figure **??** and the family of transition probabilities is provided in Table 4.2 below. It is easy also to compute the transition probabilities associated to

Contexts(w)	**p(0|w)**	**p(1|w)**	**p(2|w)**
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
01	0	0	1
20	ϵ	$1-\epsilon$	0
10	ϵ	$1-\epsilon$	0
000	0	0	1
100	0	0	1
200	ϵ	$1-\epsilon$	1

Table 4.2: *Contexts and transition probabilities on the alphabet $A = \{0,1,2\}$, for the quaternary random source.*

contexts 000 and 01:

$$p(0|000) = p(0|10) = 0,\ p(1|000) = p(1|10) = 0 \text{ and } p(2|000) = p(2|10) = 1.$$

In fact, the string 000 comes from a string 101 which had the two weak beats erased, while the string 01 was also 01 in the deterministic sequence. Moreover, in both cases after the sequences 101 and 01 in the deterministic sequence there is always a strong beat 2 which is never erased.

Notice we may identify each EEG recording associated to a given realization from one of the rhythmic sources as a stochastic process $(Y_n)_{n\in\mathbb{Z}}$ in which

$$Y_n = \big(Y_n(t) : t \in [0, T]\big)$$

is the EEG signal correspondent to the n-th acoustic stimulus generated by the stochastic source. Knowing the realization of the random source, we marked each one of the acoustic stimulus onsets according to the following convention (see figure 4.2):

- $V2$ is a strong beat,
- $V1a$ is the first weak beat (the one coming from after the symbol 2),
- $V1b$ is the second weak beat,
- $V0$ is a constitutive silence unit (present only in the Quaternary source), and
- Miss is a erased weak beat.

By doing this, we may ask, for instance, whether the law of EEG signals associated to $V1a$ is different from the law of those correspondent to $V1b$. Both $V1a$ and $V1b$ correspond to weak beats, but they are completely different from the structural point of view. In fact, in the Quarternary rhythm $V1a$ always comes from a $V2$, while $V1b$ always comes from a $V0$. In the ternary source, again $V1a$ always comes from a $V2$, but now $V1b$ comes from either a $V1a$ or Miss. In the Quaternary random source we may also ask the same question but instead of considering the two types of weak beats, $V1a$ and $V1b$, we consider the two types of silence units, $V0$ and Miss. Any difference of these laws would provide significant evidences that it is possible to retrieve structural characteristics of a stochastic source in the EEG signals driven by a sample produced by this source. In mathematical terms, this would suggest the law of Y_n is not a function only of the symbol X_n produced by the source at time n but rather it possibly depends of a portion of the past $X_n, X_{n-1}, \ldots$. This is the motivation for considering the class of hidden variable length Markov chains which we shall now define.

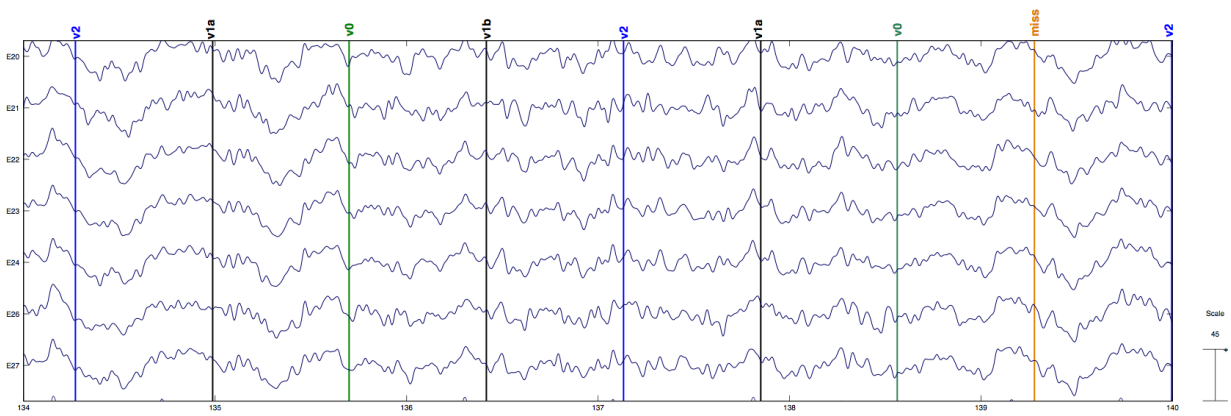


Figure 4.2: *Example of a EEG sample driven by the Quaternary source. In this figure v2 (blue) denote a strong beat, v1a and v1b (black) the first and second weak beats, v0 (green) the constitutive silence unit and miss (orange) is how we represent a weak beat which was erased.*

Definition 5. Let the pair $(F, \mathcal{F})$ be a measurable space. The bivariate stochastic chain $(X_n, Y_n)_{n\in\mathbb{Z}}$ taking values in $A \times F$ is a *hidden variable length Markov chain* compatible with the probabilistic context tree (τ, p) and the family $(Q_w : w \in \tau)$ of transition probabilities on $(F, \mathcal{F})$, if the following conditions are satisfied,

(i) $(X_n)_{n\in\mathbb{Z}}$ is a stochastic chain with memory of variable length associated to (τ, p) and

(ii) for any $m, n \in \mathbb{Z}$ with $m \leq n$, any string $x^n_{m-\ell(\tau)+1} \in A^{n+m+\ell(\tau)}$ and any sequence $I_m, \ldots, I_n$ of $\mathcal{F}$-measurable sets,

$$\mathbb{P}\Big(Y_m \in I_m, Y_{m+1} \in I_{m+1}, \ldots, Y_n \in I_n | X^n_{m-\ell(\tau)+1} = x^n_{m-\ell(\tau)+1}\Big) = \prod_{k=m}^{n} Q_{c_\tau(x^k_{k-\ell(\tau)+1})}(I_k),$$

where $\ell(\tau)$ is the height of the context tree τ (here assumed to be finite) and $c_\tau(x^k_{k-\ell(\tau)+1})$ is the context in τ assigned to the string of symbols $x^k_{k-\ell(\tau)+1}$.

Notice that the item (ii) above implies, in particular, that conditionally on the state sequence $(X_n)_{n\in\mathbb{Z}}$, the *observable process* $(Y_n)_{n\in\mathbb{Z}}$ is a sequence of independent variables such that the distribution of Y_n given the sequence of symbols $X^n_{-\infty}$ depends only on $c_\tau(X^n_{-\infty})$.

In our case study, we may, for example, consider the set of EEG signals as the set $F = L^2([0, T])$ of all real-valued square-integrable functions in $[0, T]$ and $\mathcal{F}$ is the Borel sigma-algebra of F. For a moment let us consider we have a sample $S_1, \ldots S_n$ produce by the Quaternary source. We write $I_n(21)$ and $I_n(01)$ to denote the set of indexes m such that the context assigned to S^m_1 is 21 and 01 respectively. Thus, if we model the bivariate process $(S_n, Y_n)_{n\in\mathbb{Z}}$ as a HVLMC, we can denote EEG signals associated to $V1a$ by $\{Y^{(21)}_k, k \in I_n(21)\}$ and similarly the EEG signals associated to $V1b$ by $\{Y^{(01)}_k, k \in I_n(01)\}$. In addition, we have that the sample $\{Y^{(21)}_k, k \in I_n(21)\}$ was generated by the probability measure Q_{21}, while the other sample $\{Y^{(01)}_k, k \in I_n(01)\}$ was generated by the probability measure Q_{01}. Indeed, all sequences of symbols produced by S_n ending with 21 have context 21, while those ending with 01 have context 01, see the context tree **??**. Moreover, it must hold that $Q_{21} \neq Q_{01}$, otherwise 1 would be a context in the context tree of the Quaternary source. Thus, if the EEG signals associated to the Quaternary source really carry some structure of this source, we would expect that the sub-samples $\{Y^{(21)}_k, k \in I_n(21)\}$ and $\{Y^{(01)}_k, k \in I_n(01)\}$ are produced by different laws. Therefore, we translate the problem of retrieving structural features of a random source in the EEG signals driven by a sample produced by this source into two sample hypothesis testing problem of infinite-dimensional random variables.

Motivated by the question of how retrieve structural features from EEG signals driven by a random source we pose the following problem in a more general setting. Namely, given a finite sample (X^n_1, Y^n_1) of HVLMC associated to a probabilistic context tree (τ, p), being τ a general but finite tree, and a family $(Q_w : w \in \tau)$ of transition probabilities on $(F, \mathcal{F})$, how to estimate the context tree τ based on functional sample Y^n_1. This problem requires in particular to deal with statistical analysis of functional data. We shall solve this problem for two different cases.

The first case consists in considering a simpler scenario in which $F = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of the open sets of $\mathbb{R}$, and for each $w \in \tau$, $Q_w = \mathcal{N}(0, \sigma^2_w)$ is normal distribution with mean 0 and variance σ^2_w. Although simple we shall see that this case is similar to second case we deal with, where $F = L^2([0, T])$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of F and $w \in \tau$, Q_w is a general probability measure on $\mathcal{F}$ satisfying mild assumptions. In order to treat the second case we borrow ideas from the random projective method introduced in Cuesta-Albertos et al. (2006).

4.1.1 Observable process with Normal distribution with 0 mean and variance depending on the context

Throughout the section we work on

Assumption 4.1. $F = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ is the Borel sigma-algebra of the open sets of $\mathbb{R}$, and for each $w \in \tau$, $Q_w = \mathcal{N}(0, \sigma_w^2)$ is normal distribution with mean 0 and variance σ_w^2.

Assume we are given a sample (X_1^n, Y_1^n) of a hidden variable length Markov chain associated to a probabilistic context tree (τ, p) and family $(Q_w : w \in \tau)$ of transition probabilities on $(F, \mathcal{F})$. Let d be any integer number such that $1 \leq d < n$. For any finite string $w \in \cup_{k=1}^d A^k$, we define the set of indexes $I_n(w) = \{d \leq m \leq n : X_{m-\ell(w)+1}^m = w\}$ and we write $N_n(w)$ to denote the number of occurrences of the string w in the sample X_1^n, i.e,

$$N_n(w) = \sum_{m=d}^{n} 1\{X_{m-\ell(w)+1}^m = w\}.$$

Notice that, by definition, $N_n(w) = |I_n(w)|$. Our task is then to estimate the context tree τ based on the sample Y_1^n and the variables $I_n(w)$ for all strings $w \in \cup_{k=1}^d A^k$. For each string $w \in \cup_{k=1}^d A^k$, we write $Y^{(w)} = \{Y_k^{(w)}, k \in I_n(w)\}$ to denote the observable sub-sample induced by the string w.We need also the following assumption.

Assumption 4.2. For each $w \in \tau$, if the following set

$$\Lambda_w^\tau := \big\{s \in \tau \setminus \{w\} : \ell(s) = \ell(w), s_{-i} = w_{-i}, i = 1, \ldots, \ell(w) - 1\big\} \neq \emptyset$$

then there exists $s \in \Lambda_w^\tau$ such that $\sigma_s^2 \neq \sigma_w^2$.

Fixed a sample (X_1^n, Y_1^n), for any string $w \in \cup_{k=1}^d A^k$, with $N_n(w) > 0$, the log-Likelihood function of the sub-sample $Y^{(w)}$ induced by w, conditioning on X_1^n, is given by

$$\log L_w(Y_1^n | X_1^n) = -\frac{N_n(w)}{2}\Big[\log(2\pi) + 1 + \log(\hat{\sigma}_{w,n}^2)\Big], \tag{4.2}$$

where

$$\hat{\sigma}_{w,n}^2 = \frac{1}{N_n(w)} \sum_{k \in I_n(w)} (Y_k - \hat{\mu}_{w,n})^2 \qquad \text{and} \qquad \hat{\mu}_{w,n} = \frac{1}{N_n(w)} \sum_{k \in I_n(w)} Y_k.$$

Definition 6. We will say the irreducible tree τ is *admissible* for the sample (X_1^n), if $\ell(\tau) \leq d$, $\sum_{a \in A} N_n(wa) > 0$ for any $w \in \tau$ and for any $j = d, \ldots, n-1$ there exists a context $w \in \tau$ such that $w \preceq X_1^j$.

If τ is admissible for the sample (X_1^n), the log-Likelihood function of Y_1^n with respect to τ, given the realization X_1^n, is

$$\begin{aligned} \log L_\tau(Y_1^n \mid X_1^n) &= \sum_{w \in \tau} \log L_w(Y_1^n \mid X_1^n) \\ &= -\frac{n}{2}\big[\log(2\pi) + 1\big] - \frac{1}{2}\sum_{w \in \tau} N_n(w) \log(\hat{\sigma}_{w,n}^2). \end{aligned} \tag{4.3}$$

The next theorem is main tool to control the fluctuations of the likelihood $L_\tau(Y_1^n | X_1^n)$.

Theorem 13. *Suppose that (X_1^n, Y_1^n) is a sample of an ergodic HVLMC compatible with the probabilistic context tree (τ^*, p^*) with τ^* finite $d \geq \ell(\tau^*)$ and family $(Q_w : w \in \tau)$ of transition probabilities satisfying assumptions 4.1 and 4.2. Then, the following results hold eventually almost surely as $n \to \infty$:*

(i) Underestimation. For any admissible context tree $\tau \prec \tau^*$ there exists a positive constant $c(\tau^*, \tau)$ such that

$$\log L_{\tau^*}(Y_1^n \mid X_1^n) - \log L_\tau(Y_1^n \mid X_1^n) \geq c(\tau^*, \tau)\, n \tag{4.4}$$

(ii) Overestimation. For any admissible context trees $\tau \succ \tau'$ with $\tau' \succeq \tau^*$ there exists a positive constant $c(\tau', \tau)$ such that

$$\log L_\tau(Y_1^n \mid X_1^n) - \log L_{\tau'}(Y_1^n \mid X_1^n) \leq c(\tau, \tau')(\log n)^2. \tag{4.5}$$

In the proof of the Theorem 13 we shall use the following Lemma which can be easily checked.

Lemma 4.1. *Given a context $w \in \tau^*$ and any string $s \in \cup_{k=1}^{d-\ell(w)} A^k$, it holds true:*

$$\sigma_{sw}^2 := \lim_{n\to\infty} \hat{\sigma}_{sw,n}^2 = \sigma_w^2 \text{ and } \sigma_w^2 = \sum_{s\in A} \frac{p(sw)}{p(w)} \sigma_{sw}^2.$$

Proof of Theorem 13. (i). We shall prove for the case when τ^* differs of τ for only one branch, in one step backward in the past. Formally this means that, there exists a string $w \in \tau$ such that for each $a \in A$, if $p(aw) > 0$, then $aw \in \tau*$.

Observe that, using (4.3), the inequality (4.4) can be rewritten as

$$-\frac{1}{2}\sum_{a\in A} N_n(aw)\log(\hat{\sigma}_{aw,n}^2) + \frac{1}{2}N_n(w)\log(\hat{\sigma}_{w,n}^2) \geq c\,n.$$

By the ergodic theorem, for any string w, $N_n(w)/n \to p(w)$ almost surely as $n \to \infty$. Therefore, it is enough to show that

$$-\sum_{a\in A} p(aw)\log(\sigma_{aw}^2) + p(w)\log(\sigma_w^2) > 0. \tag{4.6}$$

But, using the Jensen inequality and the second equality of Lemma 4.1, we have

$$-\sum_{a\in A} p(aw)\log(\sigma_{aw}^2) \geq -p(w)\log\Big(\sum_{a\in A}\frac{p(aw)}{p(w)}\sigma_{aw}^2\Big) = -p(w)\sigma_w^2. \tag{4.7}$$

The equality in (4.7) holds if and only if, for all $a \in A$, $\sigma_{aw}^2 = \sum_{a\in A}\frac{p(aw)}{p(w)}\sigma_{aw}^2 = \sigma_w^2$. This implies that $w \in \tau^*$ which contradicts the fact $\tau \prec \tau^*$.

(ii). Without lost of generality, we will assume that the difference between τ and τ' is given in only one branch, in one more step on the past. That is, as before, there exists a string $w \in \tau'$ such that, for each $a \in A$ with $p(aw) > 0$, $aw \in \tau$. Now, using this, the definition of the log-Likelihood, the Lemma 4.1 and the fact that $\log x \leq x - 1$ if $x > 0$, we have that the left size of (4.5) can be bounded above by

$$-\frac{1}{2}\sum_{a\in A} N_n(aw)\log(\hat{\sigma}_{aw,n}^2) + \frac{1}{2}N_n(w)\log(\sigma_w^2) \leq \frac{1}{2}\sum_{a\in A} N_n(aw)\frac{|\sigma_{aw}^2 - \hat{\sigma}_{aw,n}^2|}{\hat{\sigma}_{aw,n}^2}. \tag{4.8}$$

We will first show that the sequence $(t_n)_{n\geq 0}$, where $t_n = C(\log n)^2$, for some constant C, satisfies

$$\mathbb{P}\big(|\sigma_{n,aw}^2 - \hat{\sigma}_{aw,n}^2| > t_n \ i.o\big) = 0. \tag{4.9}$$

Define $\tilde{\sigma}_{aw,n}^2 = \frac{1}{N_n(aw)}\sum_{k\in I_n(aw)} Y_k^2$ and observe that, writing $t = t_n$, we can upper bound the last probability by

$$\mathbb{P}\big(|\sigma_{aw}^2 - \tilde{\sigma}_{aw,n}^2| > t/2\big) + \mathbb{P}\big((\hat{\mu}_{aw,n})^2 > t/2\big). \tag{4.10}$$

We will show that the second term in (4.10) goes to 0 almost surely when n goes to infinity, but the same arguments can be applied to show that the first term in (4.10) goes to 0 as well.

Define, for $k \in I_n(aw)$, the sequence of random variables

$$Z_k^{(L)} = \begin{cases} L, & \text{if } Y_k \geq L \\ Y_k, & \text{if } -L < Y_k < L \\ -L, & \text{if } -L \leq Y_k. \end{cases}$$

Notice that, since Y_k has symmetric distribution in relation to 0, for each $k \in I_n(aw)$, it follows that $\mathbb{E}[Z_k^{(L)}] = 0$. Now, using the random variables $Z_k^{(L)}$, observe that the second term in (4.10) can be rewritten and upper bounded by

$$\begin{aligned} &\mathbb{E}\Big[\mathbb{P}\Big(\frac{1}{N_n(aw)} \sum_{k \in N_n(aw)} Y_k > \sqrt{t/2} \,\big|\, N_n(aw)\Big)\Big] \\ &\quad \leq \mathbb{E}\Big[\mathbb{P}\Big(\frac{1}{N_n(aw)} \sum_{k \in I_n(aw)} Z_k^{(L)} > \sqrt{t/2} \,\big|\, N_n(aw)\Big)\Big] + \mathbb{E}\Big[\sum_{k \in I_n(aw)} \mathbb{P}\big(Y_k \neq Z_k^{(L)}\big)\Big]. \end{aligned}$$

Now, since $Y^{(aw)}$ are i.i.d random variables, so are $Z_k^{(L)}$. Then using the Hoeffding's inequality, a large deviation inequality for normal distributed random variables and the fact that $N_n(aw) \leq n$, the last inequality can be upper bounded by

$$\mathbb{E}\Big[\exp\Big\{-\frac{tN_n(aw)}{2L}\Big\}\Big] + \frac{2n}{\sqrt{2\pi\sigma_{aw}^2}} \frac{1}{L} \exp\Big\{-\frac{L^2}{2\sigma_{aw}}\Big\}. \tag{4.11}$$

Now, take $L = (2\sigma_{aw}(2+\gamma)\log n)^{1/2}$, $t = \frac{(1+\gamma)2L\log n}{N_n(aw)}$, where $\gamma > 0$. In this case, from the expression above we get that

$$\mathbb{P}\big((\hat{\mu}_{aw,n})^2 > t/2\big) \leq \exp\Big\{-(1+\gamma)\log n\Big\} + \frac{2n}{\sqrt{2\pi\sigma_{aw}^2}} \frac{1}{L} \exp\Big\{-\frac{L^2}{2\sigma_{aw}}\Big\}$$

and therefore,

$$\sum_{n \geq 1} \mathbb{P}\big((\hat{\mu}_{aw,n})^2 > t/2\big) < \infty. \tag{4.12}$$

Applying the Borel-Cantelli lemma we deduce (4.9). Now, it remains to check the the first term of 4.10 is also summable in n. As before, for $k \in I_n(aw)$ consider the sequence of random variables

$$Z_k^{(L)} = \begin{cases} L, & \text{if } Y_k^2 - \sigma_{aw}^2 \geq L, \\ Y_k^2 - \sigma_{aw}^2, & \text{if } -L < Y_k^2 - \sigma_{aw}^2 < L, \\ -L, & \text{if } -L \leq Y_k^2 - \sigma_{aw}^2. \end{cases}$$

The expected value of $Z_k^{(L)}$ is also equal to 0. Therefore, proceeding exactly as before, we can show that, taking $t_n = \frac{(1+\gamma)2L\log n}{N_n(aw)}$ but now with $L = 2\sigma_{aw}(2+\gamma)\log n$, where $\gamma > 0$, it holds that

$$\sum_{n \geq 1} \mathbb{P}\Big(|\sigma_{aw}^2 - \tilde{\sigma}_{aw,n}^2| > t_n\Big) < \infty.$$

Using (4.12) and the inequality above we conclude that

$$\sum_{n \geq 1} \mathbb{P}\big(|\sigma_{n,aw}^2 - \hat{\sigma}_{aw,n}^2| > t_n\big) < \infty, \tag{4.13}$$

where $t_n = \frac{(1+\gamma)2L\log n}{N_n(aw)}$ and $L = 2\sigma_{aw}(2+\gamma)\log n$ with $\gamma > 0$.

Defining $C' = (1/2)\min\{\sigma_{aw}^2, a \in A\}$, by the strong consistence of the maximum likelihood

estimator, we know that

$$\mathbb{P}\Big(\bigcap_{a\in A}\{\hat{\sigma}^2_{aw} > C'\} \text{ for all } n \text{ large enough}\Big) = 1. \tag{4.14}$$

Therefore, by (4.13) and (4.14) we deduce that (4.8) is upper bounded, eventually almost surely as $n \to \infty$, by

$$C \log^2 n,$$

for some constant $C > 0$, what proves (4.5). □

The Theorem 13 give us a idea of how to estimate the context tree, since the growing of the log-Likelihood in relation to the number of contexts change at the true context tree. For any finite string $w \in \cup_{k=1}^{d-1} A^k$, we define

$$\Delta_n(w) = \frac{1}{2}\sum_{a\in A} N_n(aw) \log \frac{\hat{\sigma}^2_{w,n}}{\hat{\sigma}^2_{aw,n}}. \tag{4.15}$$

Notice that $\Delta_n(w)$ is the log-likelihood ratio statistic for testing the consistency of the sample with the context tree τ against the alternative that it is consistent with $\tau' = \tau \setminus \{w\} \cup_{a\in A} \{aw\}$.

Definition 7. For any $c > 0$, given the sample (Y_1^n, X_1^n), our *context tree estimator* is given by

$$\begin{aligned}\hat{\tau}_{n,c}(Y_1^n \mid X_1^n) = \Big\{ w \in A_1^d \ : \ & N_n(w) > 0,\ \Delta_n\big(w^{-1}_{-\ell(w)+1}\big) \geq \alpha(c,n) \text{ and} \\ & \Delta_n(sw) \leq \alpha(c,n) \text{ for all } s \in A^{d-\ell(w)} \text{ with } N_n(sw) > 0\Big\}.\end{aligned} \tag{4.16}$$

where $\alpha(c,n) = c\,(\log n)^2$.

As a consequence of the Theorem 13, we have the consistency of the estimator $\hat{\tau}_{n,c}(Y_1^n \mid X_1^n)$.

Corollary 4.1. *Suppose that (X_1^n, Y_1^n) is a sample of an ergodic HVLMC compatible with the probabilistic context tree (τ^*, p^*) with τ^* finite $d \geq \ell(\tau^*)$ and family $(Q_w : w \in \tau)$ of transition probabilities satisfying assumptions 4.1 and 4.2. Denoting $\hat{\tau}_{n,c} = \hat{\tau}_{n,c}(Y_1^n \mid X_1^n)$, it holds*

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{\tau}_{n,c} \neq \tau^*\big) = 0.$$

4.1.2 Algorithm and simulation study

In this section, we provide an algorithm to compute the estimator $\hat{\tau}_{n,c}(Y_1|X_1^n)$. Assume it is given a sample (X_1^n, Y_1^n) of an ergodic HVLMC compatible with the probabilistic context tree (τ^*, p^*) with τ^* finite $d \geq \ell(\tau^*)$ and family $(Q_w : w \in \tau)$ of transition probabilities satisfying assumptions 4.1 and 4.2. The algorithm can be described informally as follows.

(i) Start with the complete tree of depth d, that is, $\tau = \{w \in A^d : N_n(w) > 0\}$.

(ii) Choose any $w \in \tau$ and compute the statistics $\Delta_n(w^{-1}_{-\ell(w)+1})$ given in (4.15).

  (a) If the value of this statistics is less than $\alpha(c,n)$, we remove from τ all strings $aw^{-1}_{-\ell(w)+1}$, $a \in A$, and repeat step (ii) for this new tree.

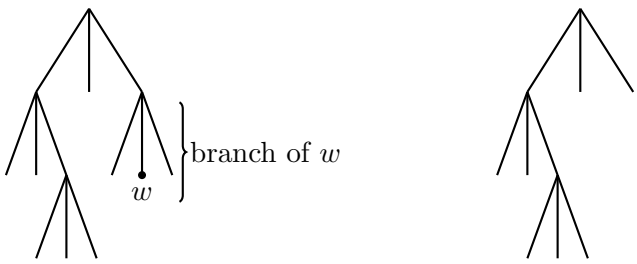
  (b) If, on the other hand, the statistics is greater than $\alpha(c,n)$, then w is a context. We then start again from the step (ii) choosing a different element of τ.

Formally, define $V_{\tau,w} = \big\{s \in \tau : \ell(s) = \ell(w) = \ell, s_{-i} = w_{-i}, i = 1, \ldots, \ell(w) - 1\big\}$. In words, $V_{\tau,w}$ is the subset of τ correspondent to the strings which are in the same branch as the string w. Now,

remember that A^* denote the set of all finite sequences of elements of A and consider the function $H_\tau : \tau \to A^* \cup \tau$ given by:

$$H_\tau(w) = \{w_{-\ell(w)+1}^{-1}\} \cup \{w' \in \tau : w' \notin V_{\tau,w}\}.$$

In words, $H_\tau(w)$ returns a context tree generated by replacing the leaf w and the leaves in its branch by one single leaf $w_{-\ell(w)+1}^{-1}$ (see Figure 4.3). The pseudo-algorithm to estimate the context tree



(a) τ (b) $H_\tau(w)$

Figure 4.3: *Example of the effect of the function $H_\tau(w)$ in a tree τ*

from a sample (X_1^n, Y_1^n) of an ergodic HVLMC $(X_n, Y_n)_{n\in\mathbb{Z}}$ is given below. In the pseudo-algorithm write $\tau[s]$ to denote the s-th element of τ.

Our goal now is to apply the proposed algorithm in the following example. We consider the ternary random source $(W_n)_{n\in\mathbb{N}}$, taking values in $A = \{0, 1, 2\}$, compatible to the context tree defined in the figure 4.1(a) whose associated family of transition probability is given in the left table of 4.3 . We then take the observable sequence $(Y_n)_{n\in\mathbb{Z}}$ whose family of transition probabilities is given in the right table of 4.3.

Waltz VLMC			
context (w)	**p(0|w)**	**p(1|w)**	**p(2|w)**
2	0.2	0.8	0
21	0.2	0.8	0
20	0.2	0.8	0
11	0	0	1
10	0	0	1
01	0	0	1
00	0	0	1

Waltz Hidden VLMC	
context (w)	$\mathbf{Q_w}$
2	$\mathcal{N}(0, 2)$
21	$\mathcal{N}(0, 4)$
20	$\mathcal{N}(0, 7)$
11	$\mathcal{N}(0, 1.7)$
10	$\mathcal{N}(0, 2.5)$
01	$\mathcal{N}(0, 7)$
00	$\mathcal{N}(0, 6)$

Table 4.3: *Transition probabilities of Waltz Hidden VLMCs simulation*

In the left side table we present for each context the transition probabilities for the next symbol on the Waltz VLMC. In the right side table we show the distribution assign to each context in the Waltz Hidden VLMC.

We consider samples of size 10 000, 20 000 and 30 000. For each sample size we simulate 100 samples. Then, for each sample we estimate the context tree by using the Algorithm **??** with the constant $c = 1/2$. In the Table 4.4, we present the proportion of time the Waltz context tree was correctly identified by the algorithm for 100 simulated sequences of sizes 10 000, 20 000 and 30 000.

Sample size	**Proportion**
2 000	0.02
5 000	0.56
10 000	1

Table 4.4: *Proportion of correct estimations of the Waltz context tree*

We highlight two important considerations about the estimation method. The first one is a natural observation about the family of normal distributions $(Q_w)_{w\in\tau}$. Since our estimator is based on empirical variances, the higher the variances in the family of normal distributions the higher the sample size required to estimate correctly the context tree.

The second one concerns the size of the sub-samples induced by the strings. Observe that, considering the transitions probability of the VLMC as in the table 4.3 we have that $\mathbb{E}[N_n(10)] = n(0.053)$ and $\mathbb{E}[N_n(00)] = n(0.013)$, while $\mathbb{E}[N_n(21)] = n(0.266)$ and $\mathbb{E}[N_n(11)] = n(0.213)$. Therefore, the sub-samples induced by 00 and 10 are quite smaller than the one induced by 21 and 11. This implies that the branch which contains the context 00 may need a greater sample size to be correctly estimated.

4.2 Hierarchical hidden context tree model

Throughout the section we work on

Assumption 4.3. $F = L^2([0, T])$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ is the Borel sigma-algebra of F and $w \in \tau$, Q_w is a general probability measure on $\mathcal{F}$ have all absolute moments finite and satisfy the Carleman's

condition, i.e, for each $n \in \mathbb{N}$ and $w \in \tau$, $m_n(w) := \int ||y||^n Q_w(dy)$ is finite and

$$\sum_{n\geq 1} m_n^{-1/n}(w) = \infty. \tag{4.17}$$

Assume we are given a sample (X_1^n, Y_1^n) of a hidden variable length Markov chain associated to a probabilistic context tree (τ, p) and family $(Q_w : w \in \tau)$ of transition probabilities on $(F, \mathcal{F})$. Our task again is to estimate the context tree τ from the sample Y_1^n and the variables $I_n(w)$, $w \in \cup_{k=1}^d A^k$, being $\ell(\tau) \leq d < n$.

To perform such estimation, we shall use a method inspired by the random projective method introduced in Cuesta-Albertos et al. (2006). In that paper, it has been shown that, under suitable conditions, a randomly chosen projection determines a distribution on a infinite-dimensional space. We use the following version of this theorem.

Theorem 14 (Cuesta-Albertos & Fraiman & Ransford (2006))**.** *Let $L^2[0,T]$ be set all square integrable functions and $< \cdot, \cdot >_{L^2}$ its correspondent inner product. Let Y_1 and Y_2 be random elements on $L^2[0,T]$ such that $\mathcal{L}(Y_1)$ satisfies that Carleman's condition, and $B = (B(t))_{t\in[0,T]}$ a Brownian Motion independent of Y_1 and Y_2. Define the random variables $R_{Y_1} = \langle Y_1, B\rangle_{L^2}$ and $R_{Y_2} = \langle Y_2, B\rangle_{L^2}$. If $\mathcal{L}(R_{Y_1}) = \mathcal{L}(R_{Y_2})$ for almost all realizations of a Brownian Motion $B = (B(t))_{t\in[0,T]}$, then $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$.*

As a byproduct we obtain the following result.

Corollary 4.2. *Let Y_1 and Y_2 be random elements of $L^2[0,T]$ such that the $\mathcal{L}(Y_1)$ satisfies the Carleman's condition. Define the real random variables*

$$V(Y_1) = \int_0^T \int_0^T Y_1(t)Y_1(s)\min\{s,t\}dsdt \quad \textit{and} \quad V(Y_2) = \int_0^T \int_0^T Y_2(t)Y_2(s)\min\{s,t\}dsdt.$$

If $\mathcal{L}(V(Y_1)) = \mathcal{L}(V(Y_2))$, then $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$.

Proof. The random variables R_{Y_1} and R_{Y_2} are Gaussian random variables with mean 0 and random variance $V(Y_1)$ and $V(Y_2)$ respectively. Indeed, since B is a standard Brownian motion, it follows that $E[R_{Y_1}|Y_1] = 0$, $E[R_{Y_1}^2|Y_1] = V(Y_1)$ and $R_{Y_1}|Y_n \sim \mathcal{N}(0, V(Y_1))$. Thus integrating the conditional $R_{Y_1}|Y_1$ with respect to the law $\mathcal{L}(Y_1)$ we get that the distribution R_{Y_1} is Gaussian random variables with mean 0 and random variance $V(Y_1)$. In short, we write $R_{Y_1} \sim N(0, V(Y_1))$ and $R_{Y_2} \sim N(0, V(Y_2))$. In particular, it follows that $\mathcal{L}(R_{Y_1})$ and $\mathcal{L}(R_{Y_2})$ do not depend on the particular realization of the Brownian Motion $B = (B(t))_{t\in[0,T]}$. Now if $\mathcal{L}(V(Y_1)) = \mathcal{L}(V(Y_2))$, then $\mathcal{L}(R_{Y_1}) = \mathcal{L}(R_{Y_2})$ and the Theorem 14 implies that $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$. □

In the previous section, to estimate the context tree τ we, for each finite string w of length at most d, compared the likelihood consistency of the sample Y_1^n with the context tree that w is considered as a context against the tree which consider as contexts all the strings aw, $a \in A$. But here, since we do not posses a closed formula for the likelihood functions we shall proceed differently. We shall, for each string $w \in \cup_{k=1}^d A^k$ and any pair $(a,b) \in A^2$, compare if the law which generated the sample $Y^{(as)} = \{Y_k, k \in I_n(as)\}$ is equal to the law which generate the sample $Y^{(bs)} = \{Y_k, k \in I_n(bs)\}$, where $s = w_{-\ell(w)+1}^{-1}$. If for all pairs $(a,b) \in A^2$, $a \neq b$ the samples $Y^{(as)}$ and $Y^{(bs)}$ have the same law, we expect that w is not a context. By Corollary 4.2, to test whether the two samples $Y^{(bw)} = \{Y_k, k \in I_n(bs)\}$ and $Y^{(as)} = \{Y_k, k \in I_n(as)\}$ come from the same distribution or not, we may test if the samples $R^{(as)} = \{R_k, k \in I_n(as)\}$ and $R^{(bs)} = \{R_k, k \in I_n(bs)\}$ come or not from the same distribution, where the variables R_k are obtained by projecting the Y_k into independent Brownian motions.

The bivariate process $(X_n, R_n)_{n\in\mathbb{Z}}$ is a "hierarchical" HVLMC. Given that $c_\tau(X_1^n) = w$, we choose Y_n according to Q_w and then we choose R_n as a normal distribution with mean 0 and variance $V(Y_n)$, that is, $R_n \sim Q_w \circ V^{-1}$ where the function V is given in the Corollary 4.2. More

precisely, the distribution of R_n for any sequence $x_{-\infty}^n$ of elements of A and measurable set I of $\mathbb{R}$, may be written as

$$\begin{aligned}\mathbb{P}\big(R_n \in I \mid X_{-\infty}^n = x_{-\infty}^n\big) &= \int_{L_2}\int_I \frac{1}{\sqrt{2\pi}V(y)} e^{-\frac{r^2}{2V(y)^2}} dr Q_{c_\tau(x_{-\infty}^n)}(dy) \\ &= \int_0^\infty \int_I \frac{1}{\sqrt{2\pi}v} e^{-\frac{r^2}{2v^2}} dr \tilde{Q}_{c_\tau(x_{-\infty}^n)}(dv) \end{aligned} \tag{4.18}$$

with $\tilde{Q}_{c_\tau(x_{-\infty}^n)} = Q_{c_\tau(x_{-\infty}^n)} \circ V^{-1}$ where for any $y \in L^2([0,T])$,

$$V(y) = \int_0^T \int_0^T \min\{s,t\}\, y(s)y(t)dsdt. \tag{4.19}$$

This is the motivation to define the following class of models:

Definition 8. The bivariate process $(X_n, R_n)_{n\in\mathbb{Z}}$ is a *hierarchical hidden variable length Markov chain* (HHVLMC) taking values in $A \times \mathbb{R}$, associated to the probabilistic context tree (τ, p) and the family $\{Q_w : w \in \tau\}$ of transition probabilities on $\mathcal{B}(\mathbb{R}_+)$, if

(i) $(X_n)_{n\in\mathbb{Z}}$ is a stochastic chain with memory of variable length associated to (τ, p) and

(ii) for any $m, n \in \mathbb{Z}$ with $m \le n$, any string $x_{m-\ell(\tau)+1}^n \in A^{n+m+\ell(\tau)}$ and any sequence $I_m, \dots, I_n$ of $\mathcal{B}(\mathbb{R})$-measurable sets,

$$\mathbb{P}\Big(R_m \in I_m, \dots, R_n \in I_n \mid X_{m-\ell(\tau)+1}^n = x_{m-\ell(\tau)+1}^n\Big) = \prod_{k=m}^n \int_0^\infty \int_{I_k} f_v(r) Q_{c_\tau(x_{k-\ell(\tau)+1}^k)}(dv), \tag{4.20}$$

where for each $r \in \mathbb{R}$ and $v \in \mathbb{R}_+$, $f_v(r) = 1/\sqrt{2\pi v^2}\exp\{-r^2/2v\}$.

We recall that our task is to estimate the context tree τ from the functional sample Y_1^n and the variables $I_n(w)$, $w \in \cup_{k=1}^d A^k$, being $\ell(\tau) \le d < n$. But as we saw in the last paragraph, we may reduce this context tree estimation from the one dimensional projected sample R_1^n and the variables $I_n(w)$. In the sequel we assume an additional hypothesis which is a generalization of the Assumption 4.2.

Assumption 4.4. Given $w = w_{-j-1}^{-1} \in \tau$ if $\Lambda_w^\tau = \{s \in \tau : s \preceq w_{-j}^{-1}\} \neq \emptyset$ then there exists $s, s' \in \Lambda_w^\tau$ such that $Q_s \neq Q_{s'}$.

The idea behind the assumption above is that, given a context $w = w_{-j-1}^{-1} \in \tau$ if for all $s, s' \in \Lambda_w^\tau$ we have $Q_s = Q_{s'}$, then we can replace all the contexts in Λ_w^τ by only one context w_{-j}^{-1} with the associated distribution Q_w without lost of information about the prediction of the process.

Generically the estimator can be described as follows. Given a string $w \in \cup_{k=1}^d A^k$, we define $s = w_{-\ell(w)+1}^{-1}$ and compare for any pair of symbols $a, b \in A$ if the samples $R^{(as)} = \{R_k, k \in I_n(as)\}$ and $R^{(bs)} = \{R_k, k \in I_n(bs)\}$ have or not common law. If for all a, b the samples $R^{(bs)} = \{R_k, k \in I_n(bs)\}$ and $R^{(as)} = \{R_k, k \in I_n(as)\}$ have the same law, then w is surely not a context. To perform this sequence of hypothesis tests we use the Kolmogorov-Simirnov (KS) statistics. Specifically, for any $a, b \in A$, $w \in \cup_{k=1}^d A^k$ and $s = w_{-\ell(w)+1}^{-1}$, the KS test statistics for the samples $R^{(bs)} = \{R_k, k \in I_n(bs)\}$ and $R^{(as)} = \{R_k, k \in I_n(as)\}$ is given by

$$D_n^{(w)}(a,b) = \sqrt{\frac{N_n(as)N_n(bs)}{N_n(as)+N_n(bs)}} \sup_{t\in\mathbb{R}} |F_n^{(as)}(t) - F_n^{(bs)}(t)|, \tag{4.21}$$

where $s = w_{-\ell(w)+1}^{-1}$, $F_n^{(as)}$ is the empirical distribution obtained from the sample $R^{(as)}$ and in the same way, $F_n^{(bs)}$ is the empirical distribution obtained from the sample $R^{(bs)}$. Therefore, in our

statistical procedure we will say that w is not a context if $D_n^w(a,b)$ is small enough for all $a,b \in A$. Otherwise we say that w is context.

The big advantage of this test is that, in the case $w \in \cup_{k=1}^d A^k$ is not a context , then the distribution of $D_n^{(w)}(a,b)$ for any $a,b \in A$ is a free distribution. Moreover, the asymptotic distribution of $D_n^{(w)}(a,b)$ is know to be for all $t > 0$

$$\lim_{n\to\infty} \mathbb{P}(D_n^{(w)}(a,b) \le t) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2t^2}. \tag{4.22}$$

Therefore, given a level α, we can find c_α such that

$$\lim_{n\to\infty} \mathbb{P}(D_n^{(w)}(a,b) > c_\alpha) = \alpha.$$

Furthermore, the test is consistent, i.e, when w is a context we have that

$$\lim_{n\to\infty} \mathbb{P}(D_n^{(w)}(a,b) > c_\alpha) = 1.$$

For any string $w \in \cup_{k=1}^d A^k$, define

$$\Delta_n(w) = \max_{a,b\in A} D_n^{(w)}(a,b).$$

Definition 9. For any $c > 0$, given the sample (X_1^n, R_1^n) our *context tree estimator* is given by

$$\begin{aligned}\hat{\tau}_{n,c}(R_1^n \mid X_1^n) = \Big\{ w \in A_1^d \ : \ & N_n(w) > 0, \Delta_n\big(w_{-\ell(w)+1}^{-1}\big) > c \text{ and} \\ & \Delta_n(sw) \le c \text{ for all } s \in A^{d-\ell(w)} \text{ with } N_n(sw) > 0 \Big\}.\end{aligned} \tag{4.23}$$

The next result we show that our estimator is a weakly consistent in the sense that for any $\alpha \in (0,1]$, there exits a constant $c_\alpha > 0$ such with probability going to 1 as the sample size n diverges it holds that $\hat{\tau}_{n,c} \succeq \tau$ and with probability going to α it holds that $\hat{\tau}_{n,c} \succ \tau$.

Proposition 4.1. *Let (X_1^n, Y_1^n) be a sample produced by HHVLMC associated to the probabilistic context tree (τ^*, p^*) and family $(Q_w : w \in \tau^*)$ of transition probabilities on $\mathcal{B}(\mathbb{R}_+)$ satisfying the Assumption 4.4. For any $\alpha \in (0,1]$, there exists a constant $c_\alpha > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}(\hat{\tau}_{n,c_\alpha} \succeq \tau^*) = 1 \ \textit{and} \ \lim_{n\to\infty} \mathbb{P}(\hat{\tau}_{n,c_\alpha} \succ \tau^*) = \alpha.$$

Proof. Indeed, notice that

$$\mathbb{P}\big(\hat{\tau}_{n,c_\alpha} \prec \tau^*\big) \le \sum_{w\in\tau^*} \mathbb{P}\Big(\Delta_n(w_{\ell(w)+1}^{-1}) \le c_\alpha\Big)$$

Now, for any $w \in \tau^*$, by Assumption 4.4 we know that there exits $a \in A$ such that $Q_w \ne Q_{as}$ with $s = w_{-\ell(w)+1}^{-1}$. Then, by definition of $\Delta_n(w)$ it follows that

$$\mathbb{P}\Big(\Delta_n(s) \le c_\alpha\Big) \le \mathbb{P}\big(D_n^{(s)}(w_{-\ell(w)}, a) \le c_\alpha\big)$$

which goes to 0 as $n \to \infty$ because of the consistence of the Kolmogorov-Smirnov test.

Now assume that there exits $w' \succ w$ and $w \in \tau^*$ such that $w' \in \hat{\tau}_{n,c_\alpha}$. Notice that defining $s = (w')_{\ell(w')+1}^{-1}$, we have

$$\mathbb{P}\big(\Delta_n(s) > c_\alpha\big) \le \sum_{a,b\in A: a\ne b} \mathbb{P}\Big(D_n^{(s)}(a,b) > c_\alpha\Big). \tag{4.24}$$

Now to conclude the proof, we need to choose c_α such that the sum over all $w' \succ w$ with $w \in \tau^*$ of the term in the right-hand side of (4.24) is less than α, in the limit as $n \to \infty$. This is possible thanks to (4.22). □

4.3 Order Estimation

In this last section we address the problem of context tree estimation for the class of Hidden Context Tree models with Gaussian and Poissonin observables. The estimator we propose here does not require an upper bound on the depth of the context tree. In the main Theorem of this section, Theorem 16, the strong consistency of this estimator is proven. Our proof relies on information-theoretic-like mixture inequalities in the same spirit of [CGG09] and [Dum14].

Let $(X_n)_{n\in\mathbb{Z}}$ be a stochastic chain with memory of variable length compatible with a probabilistic context tree (τ, p). Define the process $(Z_n)_{n\in\mathbb{Z}}$ taking values in τ by $Z_n = c_\tau(X_{-\infty}^n)$. Our first result shows that if for all n, Z_{n+1} is suffix of $Z_n X_{n+1}$, then the HCT order estimation problem is equivalent to HMM order estimation problem. Following [RST96], we call the stochastic chains with memory of variable length $(X_n)_{n\in\mathbb{Z}}$ satisfying this property of *probabilistic suffix automata*. Before stating precisely this result, we give two examples. The first example shows that both rhythmic random sources Samba and Waltz satisfy the probabilistic suffix automata property. The second example provides an simple context tree τ which is not a probabilistic suffix automata.

Example 4.1 (probabilistic suffix automata)**.** Let τ be either τ_{waltz} or τ_{samba} (see figure 4.1(b)). Any stochastic chain with memory of variable length $(X_n)_{n\in\mathbb{Z}}$ compatible with (τ, p), for any family of transition probabilities p, is a probabilistic suffix automata.

Example 4.2 (non probabilistic suffix automata)**.** Let τ be the context tree $\tau = \{0, 001, 101, 11\}$. Any stochastic chain with memory of variable length $(X_n)_{n\in\mathbb{Z}}$ compatible with (τ, p), for any family of transition probabilities p, is not a probabilistic suffix automata. Indeed, if $Z_n = 0$ and $X_{n+1} = 1$, then Z_{n+1} is not a suffix of $Z_n X_{n+1}$.

In words, the next result states the following. In the case where the stochastic chain with memory of variable length $(X_n)_{n\in\mathbb{Z}}$ is associated to a probabilistic suffix automata, them the problem of estimating the number of contexts of τ is equivalent to the problem of estimating the alphabet of a stochastic Markov chain of order 1.

Proposition 4.2. *Let $(X_n, Y_n)_{n\in\mathbb{Z}}$ be a hidden stochastic chain with memory of variable length. Let $(Z_n)_{n\in\mathbb{Z}}$ be the the stochastic process defined for each $n \in \mathbb{Z}$ by $Z_n = c_\tau(X_{-\infty}^n)$. Suppose $(X_n)_{n\in\mathbb{Z}}$ is a probabilistic suffix automata, then HCT order estimation problem for $(X_n, Y_n)_{n\in\mathbb{Z}}$ is equivalent to the HMM order estimation problem for $(Z_n, Y_n)_{n\in\mathbb{Z}}$.*

Proof. The probabilistic suffix automata assumption on $(X_n)_{n\in\mathbb{Z}}$ implies immediately that $(Z_n)_{n\in\mathbb{Z}}$ is a order 1 Markov chain. Since, conditionally on $X_{-\infty}^n$, the distribution of Y_n depends only $c_\tau(X_{-\infty}^n) = W_n$, the result follows immediately. □

In the sequel, we address the HCT order estimation problem in the general setup for two classical examples of observable processes, namely, the cases for Poissonian observable distributions and Gaussian observable distributions with know variance.

Given a context tree τ, the set of transition parameters is defined as

$$\Theta_{t,\tau} = \Big\{ p \in \mathbb{R}^{|A|\times|\tau|} : p = \big(p(a|w)\big)_{a\in A, w\in\tau} \ s.t.\ p(a|w) \geq 0 \text{ and } \forall\ w \in \tau, \sum_{a\in A} p(a|w) = 1 \Big\}, \quad (4.25)$$

while the set of parameters is given by

$$\Theta_\tau = \Big\{ (p, \mu) \in \mathbb{R}^{|A|\times|\tau|} \times \mathbb{R}^{|\tau|} : p \in \Theta_{t,\tau},\ \mu = (\mu_w)_{w\in\tau} \Big\}. \quad (4.26)$$

Take any pair of parameters $\theta = (p, \mu) \in \Theta_\tau$, and define P_θ the probability distribution under which $(X_n)_{n\in\mathbb{Z}}$ is a stochastic chain with memory of variable length compatible with (τ, p) and for any $m, n \in \mathbb{Z}$ with $\infty < m \leq n < \infty$ and any sequence $I_m, \ldots, I_n$ of $\mathcal{B}(\mathbb{R}_+)$-measurable intervals,

$$\mathbb{P}\Big(Y_m \in I_m, Y_{m+1} \in I_{m+1}, \ldots, Y_n \in I_n \mid \mathcal{F}_n\Big) = \prod_{w\in\tau} \prod_{k\in I_n(w)} Q_w(I_k),$$

where for each $w \in \tau$, $I_n(w) = \{k \leq n : c_\tau(X_{-\infty}^k) = w\}$ is the set of indexes in which the observable process is chosen according to Q_w which, in the Gaussian case, is $Q_w(dy) = g_{\mu_w}(y)dy$ with

$$g_{\mu_w}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ - \frac{(y-\mu_w)^2}{2\sigma^2}\Big\}, \tag{4.27}$$

and, in the Poissonian case, is $Q_w(dk) = g_{\mu_w}(k)\gamma(dk)$, $k \in \mathbb{N}$, $\gamma = \sum_{k=0}^\infty \delta_k$ with

$$g_{\mu_w}(k) = \frac{e^{-\mu_w}\mu_w^k}{k!}. \tag{4.28}$$

In what follows, we fix a finite sample of the observable process $Y_0, \ldots, Y_n = Y_0^n$ of a hidden context tree model with true parameter $\theta^\star$ for a given context tree $\tau^\star$ associated with the stochastic chain with memory of variable length $(X_n)_{n\in\mathbb{Z}}$ such that $(X_{n-\ell(\tau^\star)+1}^n)_{n\in\mathbb{Z}}$ is a stationary and irreducible Markov chain.

Let τ be any context tree whose length is finite, i.e, $\ell = \ell(\tau) < \infty$, and consider a probability measure ν on A^ℓ. For each pair of parameters $\theta = (p, \mu) \in \Theta_\tau$, $x_{-\ell}^{-1} \in A^\ell$ and $x_0^n \in A^{n+1}$, the conditional likelihood of the sample x_0^n given the initial sequence of symbols $x_{-\ell}^{-1}$ is defined by

$$L_\theta(x_0^n | x_{-\ell}^{-1}) = L_p(x_0^n | x_{-\ell}^{-1}) = \prod_{a\in A} \prod_{w\in\tau} p(a|w)^{N_n(w)}, \tag{4.29}$$

where $N_n(w)$ is the number of occurrences of the context w in the sample $x_{-\ell}^n$. Similarly, given the string $y_0^n \in \mathcal{B}^{n+1}$ we define the conditional likelihood of the given realization of state sequence $x_{-l}^n \in A^{n+1+\ell}$ by

$$L_\theta(y_0^n | x_{-\ell}^n) = L_\mu(y_0^n | x_{-\ell}^n) = \prod_{w\in\tau} \prod_{k\in\beta_0^n(w)} g_{\mu_w}(y_k), \tag{4.30}$$

where $I_0^n(w) = \{0 \leq k \leq n : c_\tau(x_{-\ell}^k) = w\}$. Finally, the likelihood of the sample $y_0^n \in B^{n+1}$ is defined then by

$$L_\theta(y_0^n) = \sum_{x_{-l}^{-1}\in A^l} \sum_{x_0^n \in A^{n+1}} \nu(x_{-l}^{-1}) L_\theta(x_0^n | x_{-l}^{-1}) L_\theta(y_0^n | x_{-l}^n). \tag{4.31}$$

The statistical criterion we propose is motivated by the minimum description length (MDL) principle [Ris78].

Definition 10. Given the sample $y_0^n \in \mathcal{B}^{n+1}$, the estimated context tree is denoted by $\hat{\tau}_n$ and is selected through the criterion

$$\hat{\tau}_n = \underset{\tau \text{ contex tree}}{\arg\min} \Big[- \sup_{\theta\in\Theta_\tau} \log_\theta(y_0^n) + pen(n, \tau)\Big], \tag{4.32}$$

where $pen(n, \tau)$ is the penalty term which depends on n and the context tree τ.

Notice that, if we take a permutation σ of $\{1, \ldots, |A|\}$ and define the tree context tree

$$\sigma(\tau) = \{\sigma(w) : w \in \tau\},$$

where for each $w = w_{-k}^{-1} \in \tau$, $\sigma(w) = \big(\sigma(w_{-k}) \ldots \sigma(w_{-1})\big)$, then the distribution of $(Y_n)_{n\in\mathbb{Z}}$ does not change. For this reason, we need to introduce the following definition.

Definition 11. If τ and τ' are two context tree, we say that τ and τ' are equivalent, and denote it by $\tau \sim \tau'$, if there exists a permutation σ of A such that $\sigma(\tau) = \tau'$.

To give some insights of the above definition we provide two examples.

Example 4.3. The context tree $\tau = \{0, 01, 11\}$ and $\tau' = \{1, 10, 00\}$ are equivalent. It is easy to check that $\sigma(\tau) = \tau'$, if the permutation σ is such that $\sigma(1) = 0$ and $\sigma(0) = 1$.

Example 4.4. The context tree $\tau = \{0, 1\}$ and $\tau' = \{0, 01, 11\}$ are not equivalent. In fact, the equivalence property implies that equivalent trees have the same number of contexts. Since $|\tau| = 2 \neq 3 = |\tau'|$, we see that $\tau \nsim \tau'$.

Our task in to find penalty terms which ensure the strong consistency of $\hat{\tau}_n$, that is such that $P_{\theta^\star}$-eventually almost surely $\hat{\tau}_n \sim \tau^\star$. In order to do that, we shall first compare the difference of maximum likelihood of Y_0^n and a specific mixture distribution we define below.

Given a context tree τ and $\lambda > 0$, let $\pi_{\tau,\lambda} = \pi_\tau$ be a prior probability on Θ_τ such that the following conditions hold:

(i) the vectors p and μ are independent;

(ii) ν is the uniform distribution on A^ℓ, i.e, for all $x_{-\ell}^{-1} \in A^\ell$, $\nu(x_{-\ell}^{-1}) = 1/|A|^\ell$,

(iii) if $p_w = (p(a|w))_{a\in A}$, then the vectors $(p_w)_{w\in\tau}$ are independently Dirichlet $D(1/2, \ldots, 1/2)$ distributed,

(iv) the sequence $(\mu_w)_{w\in\tau}$ is i.i.d such that $\mu_w \sim \mathcal{N}(0, \lambda)$ in the Gaussian case, and $\mu_w \sim \text{Gamma}(\lambda, 1/2)$ in the Poissonian case.

The mixture measure KT_τ on $\mathcal{B}^{n+1}$ is defined then by

$$KT_\tau(y_0^n) = \int_{\Theta_\tau} L_\theta(y_0^n)\pi_\tau(d\theta). \tag{4.33}$$

The following result provides a comparison between the maximum log-likelihood $\log L_\theta(y_0^n)$ and the mixture measure define above $KT_\tau(y_0^n)$, for each $y_0^n \in \mathcal{B}^{n+1}$. In what follows, $z_{(n)}$ and $|z|_{(n)}$ are the maximum of $z_0, \ldots, z_n$ and $|z_0|, \ldots, |z_n|$, respectively.

Theorem 15. *Grant conditions (i)-(iv), for any given finite context tree τ, $n \geq 0$ and $y_0^n \in \mathcal{B}^{n+1}$, the following inequalities hold true:*

(i) define $\alpha_{n,\tau} = \frac{(|A|-1)}{2}|\tau|\log(n+1) + (\ell(\tau)+1)\log|A|$, then

$$0 \leq \sup_{\theta\in\Theta_\tau} \log L_\theta(y_0^n) - KT_\tau(y_0^n) \leq \alpha_{n,\tau} + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n \in A^\ell} \log \frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)}, \tag{4.34}$$

(ii) For Gaussian observables,

$$\sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n \in A^\ell} \log \frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)} \leq \frac{|\tau|}{2}\log\Big(1 + \frac{(n+1)\lambda^2}{|\tau|}\Big) + \frac{|z|_{(n)}^2|\tau|}{2\lambda^2}. \tag{4.35}$$

(iii) For Poissonian observables,

$$\sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n \in A^\ell} \log \frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)} \leq \frac{|\tau|}{2}\log\Big(\frac{n+1}{|\tau|}\Big) + |\tau|\lambda z_{(n)} + \frac{|\tau|}{2}(1 + \lambda - \log 2\lambda). \tag{4.36}$$

Proof. Given $y_0^n \in \mathcal{B}^{n+1}$, we have

$$\begin{aligned}\sup_{\theta\in\Theta_\tau} \log \frac{L_\theta(y_0^n)}{KT_\tau(y_0^n)} &= \ell(\tau)\log A + \sup_{\theta\in\Theta_\tau} \log \frac{\sum_{x_{-\ell}^{-1}\in A^\ell} L_\theta(y_0^n|x_{-\ell}^{-1})}{\sum_{x_{-\ell}^{-1}\in A^\ell} KT_\tau(y_0^n|x_{-\ell}^{-1})} \\ &\leq \ell(\tau)\log A + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^{-1}\in A^\ell} \log \frac{L_\theta(y_0^n|x_{-\ell}^{-1})}{KT_\tau(y_0^n|x_{-\ell}^{-1})} \end{aligned} \tag{4.37}$$

Since $L_\theta(y_0^n|x_{-\ell}^{-1}) = \sum_{x_0^n\in A^{n+1}} L_\theta(x_0^n|x_{-\ell}^{-1})L_\theta(y_0^n|x_{-\ell}^n)$, it follows by definition that

$$KT_\tau(y_0^n|x_{-\ell}^{-1}) = \sum_{x_0^n} KT_\tau(x_0^n|x_{-\ell}^{-1})KT_\tau(y_0^n|x_{-\ell}^n).$$

Thus, by (4.37), we deduce that

$$\begin{aligned}\sup_{\theta\in\Theta_\tau} \log \frac{L_\theta(y_0^n)}{KT_\tau(y_0^n)} &\leq \ell(\tau)\log A + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n\in A^{\ell+n+1}} \log \frac{L_\theta(x_0^n|x_{-\ell}^{-1})L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(x_0^n|x_{-\ell}^{-1})KT_\tau(y_0^n|x_{-\ell}^n)} \\ &\leq \ell(\tau)\log A + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n\in A^{\ell+n+1}} \log \frac{L_\theta(x_0^n|x_{-\ell}^{-1})}{KT_\tau(x_0^n|x_{-\ell}^{-1})} + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n\in A^{\ell}} \log \frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)}\end{aligned}$$

By proposition 2.4.4 of Gassiat (2011), we have

$$\sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n\in A^{\ell+n+1}} \log \frac{L_\theta(x_0^n|x_{-\ell}^{-1})}{KT_\tau(x_0^n|x_{-\ell}^{-1})} \leq \frac{(|A|-1)}{2}|\tau|\log(n+1) + \log|A|,$$

so that it follows

$$\sup_{\theta\in\Theta_\tau} \log \frac{L_\theta(y_0^n)}{KT_\tau(y_0^n)} \leq \frac{(|A|-1)}{2}|\tau|\log(n+1) + (\ell(\tau)+1)\log A + \sup_{\theta\in\Theta_\tau} \max_{x_{-\ell}^n\in A^{\ell}} \log \frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)},$$

proving (4.34). It remains prove items (ii) and (iii). We shall first prove the item (ii). In this case, since the maximum-likelihood estimator for m_w is $\hat{m}_{n,w} = N_n(w)/(n+1)$, it follows that

$$L_\theta(y_0^n|x_{-\ell}^n) \leq \frac{1}{(\sigma\sqrt{2\pi})^{n+1}} \prod_{w\in\tau} \exp\Big\{ -\frac{\sum_{k\in I_0^n(w)} z_k^2}{2\sigma^2} - \frac{N_n(w)\hat{m}_{w,n}^2}{2\sigma^2}\Big\}. \tag{4.38}$$

After some simple algebraic manipulations, we also deduce that

$$\begin{aligned}KT_\theta(y_0^n|x_{-\ell}^n) &= \frac{1}{(\sigma\sqrt{2\pi})^{n+1}} \prod_{w\in\tau} \int_{-\infty}^{\infty} \frac{1}{\lambda\sqrt{2\pi}} \exp\Big\{ -\frac{m^2}{2\lambda^2} - \frac{1}{2\sigma^2} \sum_{k\in I_0^n(w)} (z_k - m)^2\Big\} dm \\ &= \frac{1}{(\sigma\sqrt{2\pi})^{n+1}} \prod_{w\in\tau} \frac{1}{\sqrt{1+N_n(w)\frac{\lambda^2}{\sigma^2}}} \exp\Big\{ -\frac{\sum_{k\in I_0^n(w)} z_k^2}{2\sigma^2} + \frac{N_n^2(w)\hat{m}_{w,n}^2}{2\sigma^2(N_n(w)+\frac{\sigma^2}{\lambda^2})}\Big\}.\end{aligned}$$

From (4.38) and equality above, we get that

$$\frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)} \leq \prod_{w\in\tau} \sqrt{1+N_n(w)\frac{\lambda^2}{\sigma^2}} \exp\Big\{ \sum_{w\in\tau} \frac{N_n(w)\hat{m}_{w,n}^2}{2\sigma^2(1+\frac{N_n(w)\lambda^2}{\sigma^2})}\Big\}. \tag{4.39}$$

To upper bound the first term of the product above we use Jensen inequality,

$$\prod_{w\in\tau}\sqrt{1+N_n(w)\frac{\lambda^2}{\sigma^2}}\le\Big(1+\frac{(n+1)\lambda^2}{|\tau|}\Big)^{|\tau|/2},$$

and to upper bound the second term we first notice that

$$\frac{N_n(w)}{1+N_n(w)\lambda^2/\sigma^2}=\frac{1}{\sigma^2/N_n(w)+\lambda^2}\sigma^2\le\frac{\sigma^2}{\lambda^2},$$

which implies that

$$\exp\Big\{\sum_{w\in\tau}\frac{N_n(w)\hat m_{w,n}^2}{2\sigma^2(1+\frac{N_n(w)\lambda^2}{\sigma^2})}\Big\}\le\exp\{\frac{1}{2\lambda^2}\sum_{w\in\tau}\hat m_{w,n}^2\}\le\exp\{\max_{w\in\tau}\hat m_{w,n}^2\frac{|\tau|}{2\lambda^2}\}.$$

Using also that $\max_{w\in\tau}\hat m_{w,n}^2\le|z|_{(n)}^2$, we finally obtain that

$$\sup_{\theta\in\Theta_\tau}\max_{x_{-\ell}^n\in A^\ell}\log\frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\tau(y_0^n|x_{-\ell}^n)}\le\frac{|\tau|}{2}\log\Big(1+\frac{(n+1)\lambda^2}{|\tau|}\Big)+\frac{|z|_{(n)}^2|\tau|}{2\lambda^2}.$$

Thus, we have verified inequality (4.35). To prove (4.36), we start noticing that in this case

$$L_\theta(y_0^n|x_{-\ell}^n)\le\frac{1}{\prod_{k=0}^n(y_k)!}\prod_{w\in\tau}\exp\Big\{-N_n(w)\hat m_{w,n}(1-\log\hat m_{w,n},)\Big\},\tag{4.40}$$

where $\hat m_{w,n}=N_n(w)/(n+1)$ is the maximum-likelihood estimator of m_w. Writing precisely the expression of the the conditional likelihood $L_\theta(y_0^n|x_{-\ell}^n)$, we easily deduce that

$$\begin{aligned}KT_\theta(y_0^n|x_{-\ell}^n)&=\frac{1}{\prod_{k=0}^n(z_k)!}\prod_{w\in\tau}\int_0^\infty\frac{\sqrt{\lambda}}{\Gamma(1/2)}m^{(\hat m_{w,n}N_n(w)-1/2)}\exp\Big\{-m(N_n(w)+\lambda)\Big\}dm\\&=\frac{1}{\prod_{k=0}^n(z_k)!}\prod_{w\in\tau}\sqrt{\frac{\lambda}{\pi}}\frac{\Gamma(\hat m_{w,n}N_n(w)+1/2)}{(N_n(w)+\lambda)^{(N_n(w)\hat m_{w,n}+1/2)}},\end{aligned}\tag{4.41}$$

so that we get

$$\frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\theta(y_0^n|x_{-\ell}^n)}\le\prod_{w\in\tau}\sqrt{\frac{\pi}{\lambda}}\exp\Big\{\frac{-N_n(w)\hat m_{n,w}(1-\log\hat m_{n,w})+(N_n\hat m_{n,w}+1/2)\log(N_n(w)+\lambda)}{\Gamma(N_n(w)\hat m_{w,n}+1/2)}\Big\}.\tag{4.42}$$

By Robbins-Stirling approximation formula, we have that

$$\Gamma(N_n(w)\hat m_{w,n}+1/2)\ge\sqrt{2\pi}\exp\{-N_n(w)\hat m_{w,n}-1/2\}(N_n(w)\hat m_{w,n}+1/2)^{N_n(w)\hat m_{w,n}},$$

and, therefore, applying this bound to inequality (4.42) and making some computations, we conclude that

$$\frac{L_\theta(y_0^n|x_{-\ell}^n)}{KT_\theta(y_0^n|x_{-\ell}^n)}\le\prod_{w\in\tau}\sqrt{\frac{e}{2\lambda}}\exp\Big\{\frac{1}{2}\log N_n(w)+(N_n(w)\hat m_{n,w}+1/2)\log\Big(1+\frac{\lambda}{N_n(w)}\Big)\Big\}.\tag{4.43}$$

Now, applying Jensen Inequality, we have that

$$\sum_{w\in\tau}\log N_n(w)\le|\tau|\log\Big(\frac{1}{|\tau|}\sum_{w\in\tau}N_n(w)\Big)=|\tau|\log\Big(\frac{n+1}{|\tau|}\Big),$$

and using the inequality $\log z \leq z - 1$, valid for all $z > 0$, it follows that

$$\sum_{w\in\tau}(N_n(w)\hat{m}_{n,w} + 1/2)\log\Big(1 + \frac{\lambda}{N_n(w)}\Big) \leq z_{(n)}\lambda|\tau| + \frac{\lambda|\tau|}{2}.$$

Plugging these two inequalities into (4.42), we obtain the desired inequality (4.36). □

For any given $\alpha > 0$, $n \geq 0$ and context tree τ, we set the penalty term as

$$\mathrm{pen}_\alpha(n,\tau) = \sum_{k=1}^{|\tau|}\frac{(|A|-1)k+\alpha}{2}\log(n+1) + A_{n,\tau} + B_{n,\tau} + C_{n,\tau}, \tag{4.44}$$

where the terms $A_{n,\tau}$ and $B_{n,\tau}$ are given respectively by

$$A_{n,\tau} = \sum_{k=1}^{|\tau|} a_{k,\tau}, \quad B_{n,\tau} = \sum_{k=1}^{|\tau|} b_{k,\tau}$$

where for each $k \geq 1$ and context tree τ,

$$a_{k,\tau} = \frac{|\tau|}{2}\log\Big(1 + \frac{(k+1)\lambda^2}{|\tau|}\Big)$$

and in the Gaussian case,

$$b_{k,\tau} = (\ell(\tau)+1)\log|A| \text{ and } C_{n,\tau} = 5\lambda^2|\tau|(|\tau|+1)\log(n+1), \tag{4.45}$$

while in the Poissonian Case,

$$b_{k,\tau} = \ell(\tau+1)\log|A| \text{ and } C_{n,\tau} = \frac{\log(n+1)}{\sqrt{\log\log(n+1)}}, \tag{4.46}$$

We now prove that the strong consistency of the estimator $\hat{\tau}_n$.

Theorem 16. *Let $\mathrm{pen}_\alpha(n,\tau)$ be the penalty term defined in (4.44) with $\alpha > 2$, where in the Gaussian case $b_{k,\tau}$ and $C_{n,\tau}$ are given in (4.45), and in the Poissonian case $b_{k,\tau}$ and $C_{n,\tau}$ are as in (4.46). If the sample Y_0^n was generate by a hidden stochastic chain with memory of variable length with parameters $(\tau^\star,\theta^\star)$ then $\hat{\tau}_n \sim \tau^\star$, $P_{\theta^\star}$-eventually almost surely.*

Proof. We shall prove only the consistency for the Gaussian case, being the Poissonian case treated similarly. We shall start showing that $|\hat{\tau}_n| \leq |\tau^\star|$, $P_{\theta^\star}$-eventually almost surely. By Borel-Cantelli lemma, it is enough to prove that

$$\sum_{n=1}^{\infty} P_{\theta^\star}(|\hat{\tau}_n| > |\tau^\star|) < \infty.$$

Since

$$P_{\theta^\star}\Big(|\hat{\tau}_n| > |\tau^\star|\Big) \leq P_{\theta^\star}\Big(|\hat{\tau}_n| > |\tau^\star|, G_n \leq t_n\Big) + P_{\theta^\star}\Big(G_n > t_n\Big),$$

and by Lemma 3 of Gassiat (2009), $P_{\theta^\star}\Big(G_n > t_n\Big) = O\Big((n+1)^{-3/2}\Big)$, we only have to prove that

$$\sum_{n=1}^{\infty} P_{\theta^\star}\Big(|\hat{\tau}_n| > |\tau^\star|, G_n \leq t_n\Big) < \infty. \tag{4.47}$$

For that sake, take a context tree τ such that $|\tau| > |\tau^*|$ and assume that $\hat{\tau}_n = \tau$. Then, by definition,

it follows that

$$\log L_{\theta^\star}(Y_0^n) \leq \sup_{\theta\in\Theta_\tau} \log L_\theta(Y_0^n) + \mathrm{pen}_\alpha(n,\tau^\star) - \mathrm{pen}_\alpha(n,\tau). \tag{4.48}$$

The Theorem 15 implies then that the right-hand side of the inequality above is bounded (we also use that $KT_\tau(Y_0^n) \leq 1$ and $\lambda = 1/2$) by

$$\begin{aligned}
\Delta_{n,\tau} &= \alpha_{n,\tau} + \frac{|\tau|}{2}\log\Big(1+\frac{(n+1)\lambda^2}{|\tau|}\Big) + 2|Y|^2_{(n)}|\tau| + \mathrm{pen}_\alpha(n,\tau^\star) - \mathrm{pen}_\alpha(n,\tau) \\
&= -\sum_{k=|\tau^\star|+1}^{|\tau|-1}\Big[\frac{(|A|-1)}{2}\log(n+1) + a_{k,\tau} + b_{k,\tau}\Big] - \frac{\alpha}{2}(|\tau|-|\tau^\star|) \\
&\quad +(C_{n,\tau^\star} - C_{n,\tau}) + 2|Y|^2_{(n)}|\tau|.
\end{aligned}$$

Thus, defining $G_n = |Y|^2_{(n)}$ and $t_n = 5\sigma^2\log(n+1)$, we have, by the equality above, that conditionally on $\{G_n \leq t_n\}$,

$$\Delta_{n,\tau} \leq -\frac{\alpha}{2}(|\tau|-|\tau^\star|)\log(n+1), \tag{4.49}$$

so that by (4.48) and (4.49), it follows also that

$$\begin{aligned}
P_{\theta^\star}\Big(\hat\tau_n = \tau, G_n \leq t_n\Big) &\leq \int_{z_0^n\in B^{n+1}} \frac{L_{\theta^\star}(z_0^n)}{KT_\tau(z_0^n)} 1\Big\{\log\frac{L_{\theta^\star}(z_0^n)}{KT_\tau(z_0^n)}, G_n \leq t_n\Big\} KT_\tau(z_0^n)\mu(dz_0^n) \\
&\leq (n+1)^{-\frac{\alpha}{2}(|\tau|-|\tau^\star|)}.
\end{aligned}$$

Since,

$$P_{\theta^\star}\Big(|\hat\tau_n| > |\tau^\star|, G_n \leq t_n\Big) \leq \sum_{t=|\tau*|+1}^{\infty} CT(t)(n+1)^{-\frac{\alpha}{2}(t-|\tau^\star|)},$$

where $CT(t)$ is the number of context tree with exactly t contexts, and $CT(t) \leq 16^t$ (Lemma 2 of Garivier), we deduce that

$$P_{\theta^\star}\Big(|\hat\tau_n| > |\tau^\star|, G_n \leq t_n\Big) \leq O((n+1)^{-\alpha/2}),$$

which implies (4.47), if $\alpha > 2$. The part that $|\hat\tau_n| \geq |\tau^\star|$, $P_{\theta^\star}$-eventually almost surely is corollary of Theorem 2 of Leroux 1992, see also Gassiat (2009) and Dumont 2011. □

Bibliography

[Asm03] S. Asmussen. *Applied probability and queues*. Applications of mathematics. Springer, New York, Berlin, Heidelberg, 2003. 23

[Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., second edição, 1999. A Wiley-Interscience Publication.

[CAFR06] J. A. Cuesta-Albertos, R. Fraiman e T. Ransford. Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bulletin of the Brazilian Mathematical Society*, 37(4):477–501, 2006. 34

[CGG09] A. Chambaz, A. Garivier e E. Gassiat. A minimum description length approach to hidden markov models with poisson and gauss09ian emissions. application to order identification. *Journal of Statistical Planning and Inference*, 139(3):962 – 977, 2009. 48

[Dav84] M. H. A. Davis. Piecewise-deterministic markov processes: a general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B*, 46(3):353–388, 1984. 15

[DMGLP15] A. De Masi, A. Galves, E. Löcherbach e E. Presutti. Hydrodynamic limit for interacting neurons. *Journal of Statistical Physics*, 158(4):866–902, 2015. 1, 2, 3, 4, 15, 30

[DMP91] A De Masi e E. Presutti. *Mathematical methods for hydrodynamic limits*. Lecture notes in mathematics. Springer-Verlag, 1991.

[Dum14] T. Dumont. Context tree estimation in variable length hidden markov models. *Information Theory, IEEE Transactions on*, 60(6):3196–3208, June 2014. 34, 48

[FL14] N. Fournier e E. Löcherbach. On a toy model of interacting neurons. *arXiv:1410.3263*, 2014. 15, 31

[GK02] W. Gerstner e W. Kistler. *Spiking Neuron Models: An Introduction*. Cambridge University Press, New York, NY, USA, 2002. 1, 3

[GL08] A. Galves e E. Löcherbach. Stochastic chains with memory of variable length, apr 2008. 35

[GL13] A. Galves e E. Löcherbach. Infinite systems of interacting chains with memory of variable lengtha stochastic model for biological neural nets. *Journal of Statistical Physics*, 151(5):896–921, 2013. 1, 3

[GL15] A. Galves e E. Löcherbach. Modeling networks of spiking neurons as interacting processes with memory of variable length. *arXiv:1502.06446*, 2015.

[HR11] Y. Huang e R. Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, 2011. 33

[JRBB06] J. Jehee, C. Rothkopf, J. Beck e D. Ballard. Learning receptive fields using predictive feedback. *Journal of physiology, Paris*, 100(1-3):125–132, 2006. 33

[Lap07] L. Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.*, 9:620–635, 1907. 1

[Mit83] I. Mitoma. Tightness of probabilities on $c([0,1];s)$ and $d([0,1];s)$. *Ann. Probab.*, 11(4):989–999, 1983.

[MRW12] M. Thieullen M. Riedler e G. Wainrib. Limit theorems for infinite-dimensional piecewise deterministic markov processes. applications to stochastic excitable membrane models. *Electron. J. Probab.*, 17:no. 55, 1–48, 2012.

[PBML13] A. Pouget, J. Beck, Wei Ji Ma e P. Latham. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9):1170–1178, 2013. 1

[Pre08] E. Presutti. *Scaling limits in statistical mechanics and microstructures in continuum mechanics*. Theoretical and Mathematical Physics. Springer, Dordrecht, 2008.

[Rab89] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.

[RB99] R.P.N. Rao e D.H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2:79–87, 1999. 33

[Ris78] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465 – 471, 1978. 49

[Ris83] J. Rissanen. A universal data compression system. *IEEE Trans. Inf. Theor.*, 29(5):656–664, Setembro 1983. 2, 3

[RST96] Dana Ron, Yoram Singer e Naftali Tishby. The power of amnesia: Learning probabilistic automata with variable memory length. *Machine Learning*, 25(2-3):117–149, 1996. 48

[RT14] P. Robert e J. Touboul. On the dynamics of random neuronal networks. *ArXiv*, 2014. 15

[SGGW11] D. Sterratt, B. Graham, A. Gillies e D. Willshaw. *Principles of computational modelling in neuroscience*. Cambridge University Press, Cambridge, New York, 2011. 1

[Spi70] F. Spitzer. Interaction of markov processes. *Advances in Mathematics*, 5(2):246 – 290, 1970.

[Szn91] A. S. Sznitman. Topics in propagation of chaos. Em Paul-Louis Hennequin, editor, *Ecole d'Et de Probabilits de Saint-Flour XIX 1989*, volume 1464 of *Lecture Notes in Mathematics*, páginas 165–251. Springer Berlin Heidelberg, 1991.

[WCD12] C. Wacongne, J. Changeux e S. Dehaene. A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience*, 32(11):3665–3678, 2012. 33

[WLvW+11] C. Wacongne, E. Labyt, V. van Wassenhove, T. Bekinschtein, L. Naccache e S. Dehaene. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51):20754–20759, Dezembro 2011. 33

[WZF+06] Y. Wang, L. Zhou, J. Feng, J. Wang e Z. Q. Liu. Mining complex time-series data by learning markovian models. Em *Data Mining, 2006. ICDM '06. Sixth International Conference on*, páginas 1136–1140, Dec 2006. 34

.

Index