Neural dependency structures: mathematical models and statistical methods

Università degli Studi di Torino
Dipartimento di Matematica
Scuola di Dottorato in Scienza ed Alta Tecnologia
Ciclo XXVI

Neural dependency structures: mathematical models and statistical methods

Elisa Benedetto

Advisor: Prof. Laura Sacerdote

Anni Accademici: 2011–2013
Settore Scientifico-disciplinare di afferenza: MAT/06

Università degli Studi di Torino
Scuola di Dottorato in Scienza ed Alta Tecnologia

Tesi di Dottorato di Ricerca in Scienza ed Alta Tecnologia
Indirizzo: Matematica

NEURAL DEPENDENCY STRUCTURES: MATHEMATICAL MODELS AND STATISTICAL METHODS

Elisa Benedetto

Advisor: Prof. Laura Sacerdote

XXVI ciclo

To my grandfathers
Mario and Gottardo

... I miss you ...

Abstract

Neural information processing is a challenging topic. Mathematicians, physicists, biologists and computer scientists have devoted important efforts to the study of this subject since the second half of the last century. However, despite important improvements in our knowledge, we are still far from a complete comprehension of the problem.

Many experimental data show that one of the primary ingredients of neural information processing is the dependency structure between the involved variables. However many classical mathematical neural models and the associated statistical tools for their analysis are typically based on independence assumptions. Actually the independence hypothesis often makes a model simpler and mathematically tractable, but also farther from the real nature of the problem.

To improve our knowledge of the features related to dependency properties, new models should be proposed. Furthermore specific methods for the study of dependency between the variables involved should be developed. The aim of this thesis is to give a contribution to this subject. In particular we consider a two-compartment neural model. It accounts for the interaction between different parts of the nerve cell and seems to be a good compromise between mathematical tractability and an improved realism. Then we develop suitable mathematical methods for the statistical analysis of this model as well as a method to estimate the neural firing rate in the presence of dependence.

Acknowledgements

It gives me great pleasure in expressing my gratitude to all those people who have supported me and had their contributions in making this thesis possible.
Foremost, I would like to express my sincere gratitude to my advisor, Prof. Laura Sacerdote, for the continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.
I would also like to thank all the researcher of the Probability group, Roberta, Enrico, Federico, Cristina and Maria Teresa, for their constant advises and friendly support during the course of my PhD. I really enjoyed the time spent with them working on articles and during the workshops, which we attended together.
A special thank goes to my PhD comrades, Michela, Eva, Luigi, Maura, Luisa, Giovanni, Massimo, Monica, Luisa, Giorgio, Fabio, Alberto, Michele, Ubertino, Paolo, Davide e Nicola, for tolerating my outburst in difficult moments and for being the cavies for my culinary experiments. In particular I express my heart-felt gratitude to Michela, for being a very special friend, Eva, Maura and Luisa, for our funny high level movie evenings, and Luigi for having suffered the heat as I'm always cold!
Special words of thanks go to educatrici apostole dell'opera di nostra signora universale and to my neighbours, Flavia, Mariangela and Lucio, for making me feel at

home during my weeks in Torino.
Besides Torino and the academic world, I would like to thank my best friends Stefania, Davide, Chiara and Matteo, for being with me in thick and thin. I find myself lucky to have friends like them in my life.
I would like to acknowledge also the animation team and the children of the camp, for the great experiences and emotions we share every year!
I am deeply thankful to my parents for their love, support, and sacrifices:
thanks to my dad for his helpful advices and for repairing my computer when it is out of work; thanks to my mother for her words and long phone calls at the end of hard days of research.
A special thanks goes to my grandmother, Rosy, for waiting impatiently for my return home every weekend together with her little cat, Puffetta!
Last, but not certainly not the least, I would like to thank Alberto, for his love and sweetness, for all the experiences that we had along these years and for those that are waiting us.

Grazie a ...

Desidero ricordare tutti coloro che mi hanno sostenuto ed hanno contribuito nel rendere questa tesi possibile.
Innanzitutto, vorrei esprimere la mia sincera gratitudine alla Prof. Laura Sacerdote, il mio advisor, per il suo sostegno durante gli studi a la ricerca di dottorato, per la sua pazienza, motivazione, entusiasmo e immensa cultura scientifica. I suoi consigli mi ha aiutato durante tutto il periodo della ricerca e di scrittura di questa tesi.
Vorrei anche ringraziare tutti i ricercatori del gruppo di Probabilitá, Roberta, Enrico, Federico, Cristina e Maria Teresa , per i loro consigli e il sostegno amichevole durante il corso di dottorato. Ricorderó sempre con gioia il tempo trascorso con loro per scrivere articoli o durante i workshop, a cui abbiamo partecipato insieme.
Un ringraziamento speciale va ai miei compagni di dottorato, Michela, Eva, Luigi, Maura, Luisa, Giovanni, Massimo, Monica, Luisa, Giorgio, Fabio, Alberto, Michele, Ubertino, Paolo, Davide e Nicola, per aver tollerato i miei sfoghi nei momenti difficili e per essere stati le cavie per i miei esperimenti culinari. In particolare, vorrei esprimere la mia sentita gratitudine a Michela , per essere un'amica molto speciale, ad Eva e Maura, per le nostre divertenti serate di alto livello cinematografico, e a Luigi, per aver sofferto il caldo, siccome ho sempre freddo!
Speciali ringraziamenti vanno alle educatrici Apostole dell'Opera di Nostra Signora

Universale e ai miei vicini, Flavia, Mariangela e Lucio, per avermi fatto sentire a casa durante le mie settimane torinesi.
Oltre al mondo accademico e torinese, vorrei ringraziare i miei migliori amici Stefania, Davide, Chiara e Matteo, per essere con me nella buona e nella cattiva sorte. Sono molto fortunata ad avere amici come loro nella mia vita.
Vorrei ringraziare anche l'animation team ed i bambini del campo, per le grandi emozioni e le esperienze uniche che condividiamo ogni anno!
Sono profondamente grata ai miei genitori per il loro amore, il loro supporto ed i loro sacrifici:
grazie a mio padre per i suoi utili consigli e per aver riparato il mio computer ogni volta che si é rotto;
grazie a mia madre per le sue parole e le lunghe telefonate alla fine di ogni giorno di lavoro.
Un ringraziamento particolare va a mia nonna Rosy, per aspettare con impazienza il mio ritorno a casa ogni fine settimana insieme alla sua gattina, Puffetta.
Ultimo, ma per nulla il meno importante, vorrei ringraziare Alberto, per il suo amore e la sua dolcezza, per tutte le esperienze che abbiamo vissuto in questi anni e per tutte quelle che ci stanno aspettando!

Contents

List of Figures

List of Tables

Introduction

Independence assumptions are typical of many stochastic models. They are often due more to convenience than to the nature of the problem at hand. However, there are situations where neglecting dependence effects can affect the realism of a model. Indeed the introduction of statistical positive or negative dependence leads to a better understanding of the structure of multivariate distributions and multivariate models, arising in many applications.

An important scientific field, where the study of dependency structures is becoming more and more important, is computational neuroscience. It is an interdisciplinary science that links cognitive science, psychology, electrical engineering, computer science, physics and mathematics to model brain function in terms of neural information processing properties.

A neuron is an electrically excitable cell that processes and transmits information through electrochemical signals, called action potentials. They are elicited whenever the electrical voltage of a neuron reaches a characteristic voltage threshold, as a consequence of external stimulations.
The activity of neurons is frequently described by renewal point process models,

which assume the statistical independence and identical distribution of the intervals between subsequent action potentials. However, the assumption of independence must be questioned for many different types of neurons. Indeed, thanks to more sophisticated neural recording techniques, it is clear that dependencies have a fundamental role in neural information processing [35, 82].
The phenomenon of significant serial correlation of the intervals between subsequent action potentials is a common property of neurons in various systems. In the sensory periphery it is observed in the sensory ganglion receptors of a paddle fish [83] and in the ganglion cells of a goldfish retina [74]. In central parts of mammalian brain, the same serial dependencies are reported in primate sensory cortex [73] and more recently in rat cortical neurons [34, 82].

The aim of this thesis is the development of stochastic models, computational probability methods and statistical techniques to analyse the dependency structures arising in neural activity.

However all these mathematical skills to detect and study dependency structures can be generalized and applied to other fields.
For instance, modelling the complex dependency structures of financial variables, such as asset returns in massive markets, is a fundamental research problem in the financial domain. Its extreme importance is partially demonstrated in the 2007 global financial crisis. That financial imbalance was originated from the mortgage market in the United States and it quickly spread to every cell of the global financial system. If early precautionary measures are taken according to the fundamental understanding of the global financial dependency structure, some of these crises may be avoided.
Actually large scale organizations consist of interdependent units, typically linked via information technology infrastructures. Furthermore, they increasingly collaborate and interact with other organizations, due to intense global competition and complexity of modern products and services. Security risk assessment and mitigation in such large scale organizations requires analysis of complex networks of dependencies that are often only partially observable.
Similar problems arise in the fields of social networks, epidemiology, demography

and survival analysis. It is usually assumed that individuals, forming a social system, are independent. However, they share the same environment or the same load, like parents and children in family disease aggregation. Hence the behaviour of one individual may affect the other components of the system, following more or less simple interaction networks.

Thesis objective and structure

The objective of this thesis is the mathematical analysis of the dependency structures arising in neural dynamics. In the main chapters of this thesis (Chapters 3, 4, 5 and 6) we analyse a particular class of neural models, which are able to reproduce these dependency structures. Moreover we provide computational probability methods and statistical techniques for a comprehensive mathematical study of the neural information processing in presence of dependence.

The first two chapters provide the necessary background for a better understanding of the following main chapters.

Chapter 1 is devoted to the mathematical background. In particular we introduce some important dependency measures, which are used throughout the thesis. Then we shortly review some notions on stochastic processes, recalling the definitions of Gauss-Markov diffusion processes, simple point processes and first passage time of a diffusion process. The last part of the first chapter is devoted to the probability density function estimation problem, recalling some well-known estimation methods in presence of both independent and dependent sample random variables.

Chapter 2 gives an overview of the neuronal background of this thesis, describing the principal elements of neuroanatomy and neurophysiology. Then we provide a brief review of existing mathematical neural models. In particular we focus our attention on two classes of stochastic neural models: the leaky integrate-and-fire models and the two-compartment neural models.

The leaky integrate-and-fire models owe their success to their relative simplicity jointly with their reasonable ability to reproduce neural characteristics. However they completely disregard the cellular structure of a neuron, losing some important features of the neural dynamics. On the other hand, two-compartment neural models account for the neuron geometry and seem to be a good compromise between mathematical tractability and an improved realism.

Chapter 3 is devoted to a comprehensive description of the two-compartment neural model introduced in [69]. In particular we focus on a statistical analysis of new neural features that this model is able to reproduce.
The most important novelty of this model is the ability to reproduce some dependency structures which arise from particular choices of the model parameters. We use dependency measures, that are unusual in the neural literature, like the Kendall's tau [64] and the notion of copula [84]. In particular we show that an increase in external input intensity increases the strength of the observed dependencies.

Besides its physiological functioning, a neuron is an electrically excitable cell that processes and transmits information through sequences of action potentials. One of the central questions in theoretical neuroscience is how to read out the input information from a sequence of action potentials in an accurate and efficient way.

A sequence of action potentials is typically stochastic in nature, due to the variability in the input received by neurons. Hence, one of the prerequisites in studying the neural information processing is knowing the expression of the probability distribution of the intervals between subsequent action potentials. Mathematically these intervals are often modelled by the first passage time of an underling stochastic process through a particular threshold.
One of the most widely discussed problem in probability is the distribution of different first passage times that can be considered in the applications. As regard the first passage time problem of one-dimensional diffusion processes, many numerical and analytical solutions already exist, while in the fields of multivariate diffusion processes there are still many open problems.

In Chapter 4 we introduce the first passage time problem of one component of a bivariate stochastic process, which has many applications including the two-compartment neural model described in Chapter 3. We prove that the probability density function of this first passage time is the unique solution of a new integral equation and we propose a numerical algorithm for its solution. The convergence properties of this algorithm are discussed. Then, in Chapter 5 we apply this numerical algorithm to find the marginal and joint distributions of the intervals between subsequent action potentials, simulated by the two-compartment neural model of Chapter 3.

Another prerequisite to understand the neural information processing is the rate of occurrence of action potentials. Indeed a traditional coding scheme assumes that most information about the external stimuli to a neuron is contained in the proportion of action potentials per time unit.
The proportion of events of the same kind per time unit is stochastically modelled by the hazard rate functions of a simple point stochastic process.
In case of independent and identically distributed (i.i.d.) sample random variables, many parametric and non-parametric hazard rate function estimators already exist. On the contrary the hazard rate estimation problem in presence of dependence is almost totally ignored in the literature.
In Chapter 6 we provide two non-parametric estimators for the unconditional and conditional hazard rate functions of a simple point process, in presence of dependence and we prove their convergence properties. Then we apply these estimators to solve the neural coding rate estimation problem, in presence of dependent intervals between subsequent action potentials.

CHAPTER 1

Mathematical Background

1.1 Measures of dependence

Dependency structures between random variables play an important role in different fields, such as demography, economics, epidemiology, signal processing and neuroscience. For this reason tools to measure the dependence between random variables and to analyse their joint behaviour are necessary.
In the literature there are several indices of dependence. In Section 1.1.1 we revise two of them: the correlation coefficient ρ and the Kendall's τ. In Chapter 3 we use the properties of these indices to detect the presence of dependent variables in the description of a single neuron dynamics.
Copulas are mathematical objects increasingly used to describe the joint behaviour of random variables. In section 1.1.2 we introduce basic concepts on copulas, necessary for the understanding of Chapter 5 and Chapter 6, while we refer to [39, 41, 61, 84, 117, 118] for a detailed introduction to copulas.

1.1.1 Correlation coefficient ρ and Kendall's τ

The correlation coefficient ρ is a popular measure of linear dependence ([78], Ch. 4).

Definition 1.1.1. The *correlation coefficient* ρ of the random variables X and Y is the ratio between their covariance and the product of their standard deviations:

$$\rho := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))]}{\sqrt{Var(X)}\sqrt{Var(Y)}}. \tag{1.1}$$

Given a set of n observations $\{(x_i, y_i), i = 1, \ldots, n\}$ of the bivariate random variable (X, Y), a correct estimator of the covariance between X and Y is

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}. \tag{1.2}$$

Here $\bar{x}$ and $\bar{y}$ denote the sample means associated to X and Y, respectively.
An estimator $\hat{\rho}$ of the correlation coefficient (1.1) is

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}, \tag{1.3}$$

where S_x and S_y are the sample standard deviations of X and Y, respectively.

The correlation coefficient ρ is commonly used due to its simplicity and low computational cost. However, it is well known that correlation is not equivalent to dependence:

a) two independent random variables are surely uncorrelated ($\rho = 0$), as their covariance is zero. On the contrary, uncorrelated random variables are not necessarily independent. $\rho = 0$ implies independence if and only if the random variables are normally distributed.

b) the correlation coefficient ρ detects only linear dependencies (first line of Table 1.1). Hence non-linear dependence between random variables does not prevent uncorrelation.

When non-linear dependencies are involved, the Kendall's tau, defined in [64], is a more reliable measure of dependence (see [38] and Table 1.1).

X	Y	$\rho(X,Y)$	$\tau(X,Y)$
uniform on $[0,1]$	$5X$	1	1
exponential with mean 1	X^2	0.89	1
standard normal	e^X	0.77	1

Table 1.1: Comparison between the correlation coefficient ρ and the Kendall's τ between perfectly dependent random variables X and Y. The correlation coefficient ρ correctly detects the perfect correlation between the random variables only in case of linear dependence (first line).

Definition 1.1.2. The *Kendall's* τ of the random variables X and Y is the difference between the probabilities of concordance and discordance for two independent copies (X_1, Y_1) and (X_2, Y_2) of the bivariate random variable (X, Y), that is

$$\tau := \mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - \mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right]. \tag{1.4}$$

Given a sample of n observations $\{(x_i, y_i), i = 1, \ldots, n\}$ of the bivariate random variable (X, Y), the couples (x_i, y_i) and (x_j, y_j) are concordant if $(x_i - x_j)(y_i - y_j) > 0$, $i \neq j$, otherwise they are discordant.
An estimator $\hat{\tau}$ of the Kendall's τ (1.4) is

$$\hat{\tau} = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{n(n-1)/2} \tag{1.5}$$

where $n(n-1)/2$ is the total number of pairs.

Both the estimator (1.3) and (1.5) are unbiased and range between -1 and 1:

a) in case of positive dependence both the coefficients are close to 1;

b) in case of negative dependence both the coefficients are close to -1;

a) in case of independence both the coefficients are close to 0.

However the value of these estimators is strongly dependent on the shape of the dependence law between the involved random variables, as shown by Table 1.1.

1.1.2 Copulas

Definition 1.1.3. A *bivariate copula* is a function $C : [0,1]^2 \to [0,1]$ with the following properties [84]:

- $C(u,0) = C(0,u) = 0$, $\forall u \in [0,1]$;
- $C(1,u) = C(u,1) = u$, $\forall u \in [0,1]$;
- C is 2-increasing, i.e. $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$, for all $(u_1,v_1),(u_2,v_2) \in [0,1]^2$ with $u_1 \leq u_2$ and $v_1 \leq v_2$.

The importance of copulas is strongly related with the results expressed by Sklar's theorem [117]. It establishes a correspondence between joint distributions and copulas.

Theorem 1.1.4 (Sklar's Theorem)**.** *Let H be a bivariate distribution function*

$$H(x_1,x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$$

with marginal distribution functions

$$F_1(x_1) = \mathbb{P}(X_1 \leq x_1) = \int_{-\infty}^{+\infty} H(x_1,x_2)dx_2$$
$$F_2(x_2) = \mathbb{P}(X_2 \leq x_2) = \int_{-\infty}^{+\infty} H(x_1,x_2)dx_1$$

where X_1 and X_2 are generic random variables.
Then there exists a bivariate copula C such that

$$H(x_1,x_2) = C(F_1(x_1),F_2(x_2)). \tag{1.6}$$

Conversely, for any couple of univariate distribution functions $F_1(x_1) = \mathbb{P}(X_1 \leq x_1)$ and $F_2(x_2) = \mathbb{P}(X_2 \leq x_2)$ of two random variables X_1 and X_2 and any bivariate copula C, the function

$$C(F_1(x_1),F_2(x_2)) \tag{1.7}$$

is the joint distribution function of X_1 and X_2 with marginals F_1 and F_2.
Furthermore, if F_1 and F_2 are continuous, then the copula C is unique.

Remark 1.1.5. Definition 1.1.3 and Sklar's Theorem 1.1.4 can be generalized to the multivariate case with univariate marginals. We refer to [39, 61, 117, 118] for a more general analysis of copulas, as in this thesis we consider only bivariate copulas.

Copulas separate dependency properties from marginal distributions, while these two features are mixed in the joint distribution functions. Moreover copulas are invariant under increasing and continuous transformations.

There exist different families of copulas, corresponding to different dependency structures. A graphical method to classify copulas is based on their associated scatterplots. In this thesis we make use of the Gaussian copula (Chapter 5) and the independent copula (Chapter 6).

The Gaussian copula is obtained by projecting a bivariate normal distribution on the unit square $[0,1]^2$. For a given correlation coefficient ρ, the Gaussian copula is

$$C_\rho(u,v) = \phi_\rho\left(\phi_\rho^{-1}(u), \phi_\rho^{-1}(v)\right) . \tag{1.8}$$

Here ϕ_ρ is the bivariate normal distribution with mean vector zero and correlation coefficient ρ, ϕ_ρ^{-1} denotes its inverse. Note that if in (1.6) one uses a Gaussian copula and non-Gaussian marginal distribution functions, the joint distribution is not a bivariate normal distribution.
A particular copula is the independent copula:

$$\Pi(u,v) = uv. \tag{1.9}$$

It coincides with the joint distribution function of two independent uniform random variables on $[0,1]$. Therefore the scatterplot of an independent copula coincides with a sample of randomly spaced points in the unit square $[0,1]^2$, as in Figure 1.2.
In Figure 1.1 and 1.2 we compare the scatterplots of a bivariate Gaussian copula with correlation coefficient $\rho = 0.9$ and a bivariate independent copula.

The shape of a copula $C(u,v)$ can be determined from modelling arguments or can be argued from plots and confirmed through statistical tests. In this thesis we follow

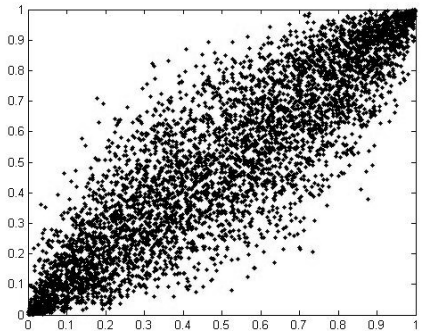


Figure 1.1: Scatterplot of a Gaussian copula ($\rho = 0.9$).

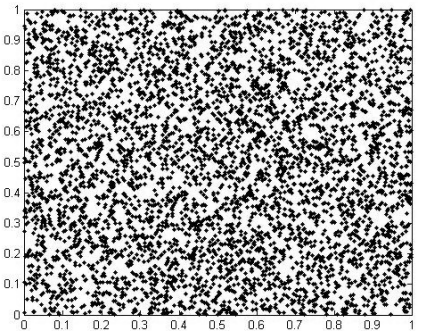


Figure 1.2: Scatterplot of an independent copula.

this last procedure, using the goodness-of-fit test proposed in [37] and revised in [42]. It is based on the comparison of the empirical copula with a parametric estimate of the copula derived under the null hypothesis.

Definition 1.1.6. The *empirical copula* of random variables X and Y is defined as

$$C_n(u, v) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(U_i \leq u, V_i \leq v), \quad (u, v) \in [0, 1]^2, \tag{1.10}$$

where $\mathbb{1}$ is the indicator function.
Given n independent copies (X_i, Y_i), $i = 1, \ldots, n$, of the random vector (X, Y), U_i and V_i denote the associated *pseudo-observations*

$$U_i = \frac{n\hat{F}_X(X_i)}{n+1}, \tag{1.11}$$

$$V_i = \frac{n\hat{F}_Y(Y_i)}{n+1}. \tag{1.12}$$

Here $\hat{F}_X$ and $\hat{F}_Y$ are the empirical distribution functions of X and Y, respectively,

$$\hat{F}_X(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \leq t), \tag{1.13}$$

$$\hat{F}_Y(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(y_i \leq t), \tag{1.14}$$

where x_i and y_i, $i = 1, \dots n$, are n observations of the random variables X and Y, respectively.

Copula goodness-of-fit tests verify whether a copula C belongs to an assumed class of copulas $\mathcal{C}_0 = \{C_\theta : \theta \in \mathcal{O}\}$, where $\mathcal{O}$ is an open subset of $\mathbb{R}^p$, $p \geq 1$, and θ is an unknown parameter of the copula.
A natural testing procedure consists in studying the "distance" between the empirical copula C_n and a parametric estimation C_{θ_n} of the assumed copula under the null hypothesis $H_0 : C \in \mathcal{C}_0$. In [37] goodness-of fit tests based on the empirical distance

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n}). \tag{1.15}$$

are briefly considered . Their implementation is examined in details in [42], where the authors consider, as test statistic, a rank-based version of the Cramér-Von Mises statistic [19, 128]:

$$S_n = \int_0^1 \int_0^1 \mathbb{C}_n(u, v)^2 dC_n(u, v). \tag{1.16}$$

Large values of this statistic lead to reject the null hypothesis H_0.

The test p-values can be deduced from the limiting distribution of S_n, which depends on the asymptotic behaviour of $\mathbb{C}_n$. The convergence of the latter follows from appropriate regularity conditions on the parametric family $\mathcal{C}_0$ and the sequence $\{\theta_n\}$ of estimators. In practice, the asymptotic distribution of S_n cannot be tabulated and approximate p-values can only be obtained via particular adapted Monte Carlo methods or specific parametric bootstrap procedures.
This copula goodness-of-fit test is consistent, i.e. if the null hypothesis is false it is rejected with probability 1 as $n \to \infty$ (for the complete proof see [42]).

Remark 1.1.7. A possible improvement of this copula goodness-of-fit test concerns a better definition of the empirical copula (1.10).
Despite their good asymptotic properties, the empirical distribution functions (1.13) and (1.14) exhibit a slow rate of convergence, see for example [94]. Therefore, we may need to resort to more reliable techniques to estimate the required distribution functions. In [52] alternative estimators, like kernel-based and nearest neighbour estimators, are proposed and their performances compared.

1.2 Stochastic processes

The largest part of modelling instances, from physics to chemistry, engineering, economics, biology or neuroscience, deals with random phenomena evolving in time. Stochastic processes are the mathematical tools, devoted to model these instances. Indeed they concern sequences of random events governed by probabilistic laws [62].

Definition 1.2.1. A *d-dimensional stochastic process* is a collection of multivariate random variables $\{\mathbf{X}(t), t \in T\}$, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in a subset of $\mathbb{R}^d$. The set T is a subset of $\mathbb{R}$, called *parameter set*.

Any stochastic process is formally function of two variables: $\omega \in \Omega$ and $t \in T$. If we fix $\omega \in \Omega$, we obtain a function of t, called trajectory of the process. Otherwise if we fix $t \in T$ we obtain a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$.

Besides their general definition, stochastic processes can be classified according to their specific stochastic properties, as for instance Markov, Gaussian or Martingale properties. Here we revise two families of stochastic processes, applied in the following sections, while we refer to [30, 62, 63, 72, 102, 107, 127] for a complete introduction to stochastic processes.

In Section 1.2.1 we introduce the class of Gauss-Markov diffusion processes. They are used in Section 2.2.2, to define a two-compartment model of a neuron. In Section 1.2.2 we describe the class of simple point processes. Their fundamental properties are applied in Chapter 6 to derive strongly consistent estimators of the hazard rate functions of a sequence of events.

In some applications of stochastic processes (e.g. biology, neuroscience and economics), the process' first attainment of a boundary is the random variable of interest. It is called first passage time (FPT). In Section 1.2.3 we formally defined the FPT and the FPT problem, i.e. the study of the FPT distribution. Moreover in this section we introduce a particular FPT problem, which is solved numerically in Chapter 4 for the class of Gauss-Markov processes.

1.2.1 Gauss-Markov processes and their properties

An important class of stochastic processes is the class of diffusion stochastic processes [5, 63, 95, 102].

Definition 1.2.2. A d-dimensional stochastic process $\{\mathbf{X}(t), t \in T\}$ on $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in $E \in \mathbb{R}^d$, is a *d-dimensional diffusion process* if its trajectories are continuous with probability 1 and it satisfies the Markov property:

$$\mathbb{P}(\mathbf{X}(t) \in B|\mathcal{F}_\tau) = \mathbb{P}(\mathbf{X}(t) \in B|\mathbf{X}(\tau)),\ \forall t > \tau,\ t, \tau \in T, \tag{1.17}$$

where B is a Borel subset of $\mathbb{R}^d$ and $\mathcal{F}_\tau$ is the process history until τ.

A d-dimensional diffusion process is completely characterized by a d-dimensional vector $\mu(\mathbf{x}, t) = (\mu_1(\mathbf{x}, t), \mu_2(\mathbf{x}, t), \ldots, \mu_d(\mathbf{x}, t))$, called drift, and a $d \times d$ positive defined matrix $Q(\mathbf{x}, t) = [\sigma_{ij}(\mathbf{x}, t)]_{i,j=1,2,\ldots d}$, called diffusion matrix [63]. Their components are defined by the following limits, for every $i, j = 1, 2, \ldots, d$,

$$\mu_i(\mathbf{x}, t) = \lim_{h \to 0^+} \frac{1}{h} \mathbb{E}\left[X_i(t + h) - X_i(t)|\mathbf{X}(t) = \mathbf{x}\right], \tag{1.18}$$

$$\sigma_{ij}(\mathbf{x}, t) = \lim_{h \to 0^+} \frac{1}{h} \mathbb{E}\left[(X_i(t + h) - X_i(t))(X_j(t + h) - X_j(t))|\mathbf{X}(t) = \mathbf{x}\right], \tag{1.19}$$

where X_i denotes the i-th component of the d-dimensional process $\mathbf{X}$.

Thanks to the Markov property, the transition density function

$$f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right) = \frac{\partial}{\partial \mathbf{x}} \mathbb{P}(\mathbf{X}(t) < \mathbf{x}|\mathbf{X}(t_0) = \mathbf{y}), \quad t > t_0, \tag{1.20}$$

of a diffusion process $\mathbf{X}$ satisfies the following key equations [63]:

- the Chapman-Kolmogorov equation

$$f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right) = \int_{\mathbb{R}^d} f\left(\mathbf{x}, t \mid \mathbf{z}, u\right) f\left(\mathbf{z}, u \mid \mathbf{y}, t_0\right) d\mathbf{z}\ , \tag{1.21}$$

- the Kolmogorov backward equation

$$\frac{\partial f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)}{\partial t_0} + \sum_{i=1}^{d} \mu_i(\mathbf{y}, t_0) \frac{\partial f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_{ij}(\mathbf{y}, t_0) \frac{\partial^2 f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)}{\partial y_i \partial y_j} = 0 \,, \tag{1.22}$$

- the Kolmogorov forward equation

$$\frac{\partial f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left(\mu_i(\mathbf{x}, t) f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)\right) - \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i x_j} \left(\sigma_{ij}(\mathbf{x}, t) f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right)\right) = 0 \,, \tag{1.23}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_d)$.

Any d-dimensional diffusion process $\{\mathbf{X}(t), t \in T\}$ is solution of a particular stochastic differential equation:

$$\begin{cases} d\mathbf{X}(t) = \mathbf{m}(\mathbf{X}(t), t)dt + \mathbf{G}(\mathbf{X}(t), t)d\mathbf{B}(t), & t \geq t_0 \\ \mathbf{X}(t_0) = \mathbf{y} \end{cases} \tag{1.24}$$

where $\mathbf{B}(t)$ is a d-dimensional standard Brownian motion. Here $\mathbf{m}(\mathbf{x}, t)$ and $\mathbf{G}(\mathbf{x}, t)$ are respectively a d-dimensional vector and a $d \times d$ matrix, whose components are measurable Lipschitz functions with respect to $\mathbf{x}$. In particular $\mathbf{m}(\mathbf{x}, t)$ and $\mathbf{G}(\mathbf{x}, t)\mathbf{G}'(\mathbf{x}, t)$ represent respectively the drift and the diffusion matrix of the diffusion process solution of (1.24). Here the superscript $'$ denotes the transpose of a matrix. Therefore, another way to characterize a diffusion process is through the stochastic differential equation of which it is solution.

Definition 1.2.3. A *d-dimensional diffusion process with linear drift*, $\mathbf{X}$, is the solution of a linear (in the narrow sense) stochastic differential equation [5]

$$\begin{cases} d\mathbf{X}(t) = \left[\mathbf{A}(t)\mathbf{X}(t) + \mathbf{M}(t)\right] dt + \mathbf{\Sigma}(t)d\mathbf{B}(t), & t \geq t_0 \\ \mathbf{X}(t_0) = \mathbf{y} \end{cases} \tag{1.25}$$

where $\mathbf{A}(t)$ and $\mathbf{\Sigma}(t)$ are $d \times d$ matrices, $\mathbf{M}(t)$ is a d-dimensional vector and $\mathbf{B}(t)$ is a d-dimensional standard Brownian motion.

The solution of (1.25), corresponding to an initial value $\mathbf{y}$ at time t_0, is

$$\mathbf{X}(t) = \phi(t, t_0) \left[\mathbf{y} + \int_{t_0}^{t} \phi(u, t_0)^{-1} \mathbf{M}(u) du + \int_{t_0}^{t} \phi(u, t_0)^{-1} \mathbf{G}(u) d\mathbf{B}(u) \right], \quad (1.26)$$

where $\phi(t, t_0)$ is the solution of the homogeneous matrix equation

$$\frac{d}{dt}\phi(t, t_0) = \mathbf{A}(t)\phi(t, t_0) \text{ with } \phi(t_0, t_0) = \mathbf{I}. \quad (1.27)$$

For $t \geq 0$, the diffusion process has a D-dimensional distribution with mean vector

$$\mathbf{m}(t\,|\mathbf{y}, t_0) := \mathbb{E}(\mathbf{X}(t)|\mathbf{X}(t_0) = \mathbf{y}) = \phi(t, t_0) \left[\mathbf{y} + \int_{t_0}^{t} \phi(u, t_0)^{-1} \mathbf{M}(u) du \right] \quad (1.28)$$

and $D \times D$ conditional covariance matrix

$$\mathbf{Q}(t\,|\mathbf{y}, t_0) = \phi(t, t_0) \left[\int_{t_0}^{t} \phi(u, t_0)^{-1} \mathbf{G}(u) \mathbf{G}(u)' (\phi(u, t_0)^{-1})' du \right] \phi(t, t_0)', \quad (1.29)$$

where the superscript $'$ denotes the transpose of a matrix.

In the autonomous case, $\mathbf{A}(t) = \mathbf{A}$, $\mathbf{M}(t) = \mathbf{M}$ and $\mathbf{\Sigma}(t) = \mathbf{\Sigma}$, expressions (1.28) and (1.29) are simplified:

$$\mathbf{m}(t\,|\mathbf{y}, t_0) = e^{\mathbf{A}(t-t_0)} \left[\mathbf{y} + \int_{t_0}^{t} e^{-\mathbf{A}(u-t_0)} \mathbf{M} du \right] \quad (1.30)$$

$$\begin{aligned} \mathbf{Q}(t\,|\mathbf{y}, t_0) &= e^{\mathbf{A}(t-t_0)} \left[\int_{t_0}^{t} e^{-\mathbf{A}(u-t_0)} \mathbf{G}\mathbf{G}' e^{-\mathbf{A}'(u-t_0)} du \right] e^{\mathbf{A}'(u-t_0)} \quad (1.31) \\ &= \int_{t_0}^{t} e^{\mathbf{A}(t-u)} \mathbf{G}\mathbf{G}' e^{\mathbf{A}'(t-u)} du. \end{aligned}$$

Definition 1.2.4. When the initial condition $\mathbf{y}$ is constant or Gaussian, the solution of (1.25) is a Gaussian process, frequently known as *Gauss-Markov diffusion process* [5].

Examples of Gauss-Markov diffusion processes are the Integrated Brownian Motion (IBM), the Integrated Ornstein Uhlenbeck Process (IOU). The underlying process of the two-compartment neural model, described in Section 2.2.2, is also a Gauss-Markov diffusion process.

If $\det \mathbf{Q}(t \mid \mathbf{y}, t_0) \neq 0$ for each t, the transition probability density function (1.20) of any two-dimensional Gauss-Markov diffusion process is

$$f(\mathbf{x}, t \mid \mathbf{y}, t_0) = \frac{\exp\left\{-\frac{1}{2}\left[\mathbf{x} - \mathbf{m}(t \,|\mathbf{y}, t_0)\right]' \mathbf{Q}(t \,|\mathbf{y}, t_0)^{-1} \left[\mathbf{x} - \mathbf{m}(t \,|\mathbf{y}, t_0)\right]\right\}}{2\pi\sqrt{\det \mathbf{Q}(t \mid \mathbf{y}, t_0)}} \tag{1.32}$$

and verifies the Chapman-Kolmogorov equation (1.21) [95].

1.2.2 Simple point processes and their properties

Definition 1.2.5. A *point process on the real line* is a stochastic process $\{N(t), t \geq 0\}$, that counts the number of events on $[0, t]$ [18, 21]. Therefore:

- $N(t) \geq 0$;
- $N(t)$ is an integer.
- If $s \leq t$ then $N(s) \leq N(t)$;
- If $s \leq t$ then $N(t) - N(s)$ is the number of events in $(s, t]$.

Definition 1.2.6. A stochastic point process on the real line N is called *simple* if at any instant at most one single event occurs with probability 1:

$$\mathbb{P}\left\{\lim_{\Delta t \to 0} [N(t + \Delta t) - N(t)] = 0 \text{ or } 1\right\} = 1, \quad \forall t \geq 0. \tag{1.33}$$

In Section 6 we consider a simple point process $N = \{N(t), t \geq 0\}$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, adapted to the natural filtration $\mathcal{N}_t = \sigma\{N(\tau), \tau \leq t\}$ and observed on a fixed time interval $[0, L]$, $0 < L < \infty$.
We denote by $l_1, l_2, \ldots, l_{N(L)}$ the ordered set of event instants in $[0, L]$. Then the inter-event interval process $\{T_i = l_i - l_{i-1},\ i \geq 1 \text{ and } l_0 = 0\}$ is determined by well-defined non-negative random variables.

To study a simple point process with inter-event intervals T_i, $i \geq 1$, we need the following additional filtrations:

$$\mathcal{F}_i = \sigma\{T_j, j = 1, 2, \ldots, i\}, \quad i \geq 1, \tag{1.34a}$$

$$\mathcal{G}_i = \sigma\{(T_j, T_{j+1}), j = 1, 2 \ldots, i\}, \quad i \geq 1. \tag{1.34b}$$

Note that (1.34a) represents the past history of the inter-event interval process, while (1.34b) is the joint past history of an inter-event interval and its subsequent.

Denote by $f_1(t)$ the unconditional density function of T_1 and by $f_i(t|\mathcal{F}_{i-1})$ the conditional density function of T_i, $i \geq 2$, given the history $\mathcal{F}_{i-1}$. In Chapter 6 we express these densities in terms of the associated hazard rate functions:

$$h_1(t) = -\frac{d}{dt} \ln [S_1(t)] = \frac{f_1(t)}{S_1(t)} , \tag{1.35a}$$

$$h_i(t|\mathcal{F}_{i-1}) = -\frac{d}{dt} \ln [S_i(t|\mathcal{F}_{i-1})] = \frac{f_i(t|\mathcal{F}_{i-1})}{S_i(t|\mathcal{F}_{i-1})}, \quad i \geq 2, \tag{1.35b}$$

where $S_1(t) = 1 - \int_0^t f_1(s)ds$ and $S_i(t|\mathcal{F}_{i-1}) = 1 - \int_0^t f_i(s|\mathcal{F}_{i-1})ds$ are the corresponding survival functions.
Therefore

$$f_1(t) = h_1(t)S_1(t) = h_1(t) \exp\left(-\int_0^t h_1(u)du\right), \tag{1.36a}$$

$$\begin{aligned} f_i(t|\mathcal{F}_{i-1}) &= h_i(t|\mathcal{F}_{i-1})S_i(t|\mathcal{F}_{i-1}) \\ &= h_i(t|\mathcal{F}_{i-1}) \exp\left(-\int_0^t h_i(u|\mathcal{F}_{i-1})du\right). \end{aligned} \tag{1.36b}$$

Another important function in the theory of simple point processes is the conditional intensity function. It measures the proportion of events per time unit, conditioned on the past history of the process. In this thesis we express the conditional intensity function in terms of the inter-event interval hazard rate functions (1.35a) and (1.35b), as in [21] (Chapter 7).

Definition 1.2.7. The *conditional intensity function* of a stochastic simple point process N is the following piecewise-defined function

$$\lambda^*(t) := \begin{cases} h_1(t), & (0 < t \leq l_1), \\ h_i(t - l_{i-1}|\mathcal{F}_{i-1}), & (l_{i-1} < t \leq l_i,\ i \geq 2). \end{cases} \tag{1.37}$$

Remark 1.2.8. Notice that $\lambda^*(t)$ is a function of the point process history up to time t. Hence it is itself a stochastic process $\lambda^*(\cdot, \omega)$. It depends on a random event ω through the realization of the inter-event interval process $\{T_i(\omega),\, i \geq 1\}$.

A well-known result on the conditional intensity function of a simple point process is the so called time-rescaling theorem ([21], Ch. 7).

Theorem 1.2.9 (Time-rescaling theorem)**.** *Let N be a simple point process, with bounded and strictly positive conditional intensity function $\lambda^*(t)$. Define $\Lambda^*(t)$ as the point-wise integral*

$$\Lambda^*(t) = \int_0^t \lambda^*(u)du\ . \tag{1.38}$$

Then, under the random time transformation

$$t \mapsto \Lambda^*(t)\ , \tag{1.39}$$

the transformed process $\tilde{N}(t) = N(\Lambda^{*-1}(t))$ is a unit-rate Poisson process.

Remark 1.2.10. The transformed inter-event intervals of $\tilde{N}(t)$ are

$$\tilde{T}_i = \Lambda^*(l_i) - \Lambda^*(l_{i-1}) = \int_{l_{i-1}}^{l_i} \lambda^*(u)du,$$

where $l_0 = 0$ and l_i, $i \geq 1$, are the event instants of the original process. According to Theorem 1.2.9, they are i.i.d. exponential random variables with mean 1.

1.2.3 FPT problems for multivariate stochastic processes

Definition 1.2.11. The *FPT* of a stochastic process is the random time taken by the process to reach an assigned threshold for the first time. The analysis of the FPT distribution is usually called, *FPT problem.*

The FPT problem arises in many different fields, like neuroscience, reliability theory, finance, and epidemiology ([93, 95, 119] and examples cited therein).
In some instances, the random variable of interest is the FPT of one of the components of a multivariate stochastic process, as for the two-compartment neural model defined in Section 2.2.2. In Chapter 4 we solve this problem numerically for a bivariate Gauss-Markov diffusion process. Here we introduce the necessary notations.

Let $\mathbf{X}(t) = (X_1(t), X_2(t))$, $t \geq t_0$, be a two-dimensional Gauss-Markov diffusion process originated in $\mathbf{y} = (y_1, y_2)$ at time t_0. The FPT of the second component through a boundary $S > y_2$ is:

$$T = \inf \{t \geq t_0 : X_2(t) \geq S\}. \tag{1.40}$$

Its probability density function is

$$g(t \,|\mathbf{y}, t_0) = \frac{\partial}{\partial t}\mathbb{P}(T < t \,|\mathbf{X}(t_0) = \mathbf{y}). \tag{1.41}$$

Another quantity of interest is the probability density function of the bivariate random variable $(X_1(T), T)$:

$$g_c((z, S), t \,|\mathbf{y}, t_0) = \frac{\partial^2}{\partial z \partial t}\mathbb{P}(X_1(T) < z, T < t \,|\mathbf{X}(t_0) = \mathbf{y}), \tag{1.42}$$
$$z \in \mathbb{R}, \quad t \in [t_0, \infty], \quad S > y_2.$$

The following theorem links the transition probability density function (1.32) of a bivariate Gauss-Markov diffusion process $\mathbf{X}$ with the joint probability density function (1.42).

Theorem 1.2.12. *For $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, with $x_2 > S$,*

$$\begin{aligned} &\mathbb{P}(\mathbf{X}(t) > \mathbf{x} | \mathbf{X}(t_0) = \mathbf{y}) \\ &= \int_{t_0}^{t} d\vartheta \int_{-\infty}^{+\infty} g_c((z, S), \vartheta \mid \mathbf{y}, t_0)\, \mathbb{P}(\mathbf{X}(t) > \mathbf{x} | X_1(\vartheta) = z, X_2(\vartheta) = S)\, dz \end{aligned} \tag{1.43}$$

and

$$f(\mathbf{x}, t \mid \mathbf{y}, t_0) = \int_{t_0}^{t} d\vartheta \int_{-\infty}^{+\infty} g_c((z, S), \vartheta \mid \mathbf{y}, t_0)\, f(\mathbf{x}, t \mid (z, S), \vartheta)\, dz. \tag{1.44}$$

Proof. Equation (1.43) is a consequence of the strong Markov property, as explained in the following.
Let $\varphi : \mathbb{R} \times (S, \infty) \to \mathbb{R}$ be a bounded, Borel measurable function and let $\mathcal{F}_T$ be the σ-algebra generated by the process $\mathbf{X}(t)$ up to the random time T. We get

$$\begin{aligned} \mathbb{E}[\varphi(\mathbf{X}(t))|\mathbf{X}(t_0) = \mathbf{y}] &= \mathbb{E}[\mathbb{E}[\varphi(\mathbf{X}(t))|\mathcal{F}_T; \mathbf{X}(t_0) = \mathbf{y}]] \\ &= \mathbb{E}[\mathbb{E}[\varphi(\mathbf{X}(t))|\mathbf{X}(T)]] \\ &= \int_{t_0}^{t} d\vartheta \int_{-\infty}^{+\infty} \mathbb{E}[\varphi(\mathbf{X}(t))|\mathbf{X}(\vartheta) = (z, S)]\ g_c((z, S), \vartheta|\mathbf{y}, t_0)dz \end{aligned} \tag{1.45}$$

where the first equality uses the double expectation theorem while the second one uses the strong Markov property. Here expectations are with respect to the probability measure induced by the random variable $\mathbf{X}(t)$.
For $\varphi(\mathbf{y}) = \mathbb{1}_{\{x_1,\infty\}\times\{x_2,\infty\}}(\mathbf{y})$ we get (1.43). Finally, writing the conditional probability $\mathbb{P}\left(\mathbf{X}(t) > \mathbf{x}|X_1(\vartheta) = z, X_2(\vartheta) = S\right)$ as a double integral, by changing the order of integration and differentiating (1.43) with respect to x_1 and x_2, we get (1.44). □

Remark 1.2.13. Equation (1.44) is introduced in [50] and and we prove it in [6].

1.3 Probability density function estimation

The probability density function is a fundamental concept in probability theory and statistics, as it gives a natural description of the distribution of continuous random variables.

Definition 1.3.1. Let X be a continuous random variable. Then the *probability density function* of X is a function $f : \mathbb{R} \to \mathbb{R}$ such that:

a) $f(x) \geq 0$, $\forall x \in \mathbb{R}$;

b) $\int_{-\infty}^{+\infty} f(x)dx = 1$;

c) $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$, $\forall a, b \in \mathbb{R}$, $a < b$;

In many instances we have a set of observed data, assumed to be a sample from an unknown probability density function. The probability density function estimation problem consists in the construction of an estimate of a probability density function from sample data.

There is a wide variety of probability density function estimators for i.i.d. sample random variables [90, 104, 105, 116]. In Section 1.3.1 we provide a brief summary of the main methods for probability density function estimation.
However the i.i.d. assumption is too strong for many applications [35, 82]. Indeed in many instances the dependence is very important, as it captures important features of the system under study. In Section 1.3.2 we introduce two examples of probability density function estimators in the presence of dependence. They are used in Chapter 6, to provide uniform strongly consistent estimators for a point process hazard rate functions (1.35a) and (1.35b).

1.3.1 Estimators under the classical i.i.d hypothesis

There exist two different approaches to the probability density function estimation problem. One approach is parametric. Assume that data are sampled from a known probability distribution, with probability density function depending on an unknown parameter. Then the underling probability density function could be estimated by finding an estimator of the unknown parameter. The second approach is non-parametric. It does not specify the underlying form of the distribution and it estimates the probability density function directly from sample data.

Parametric estimation

The most popular method to estimate the unknown parameters of a probability density function is maximum likelihood estimation ([78], Ch. 7).
Suppose there is a sample $(x_1, x_2, \ldots, x_n)$ of n i.i.d. observations, coming from a distribution with probability density function $f(\cdot) = f(\cdot\,;\theta)$ depending on an unknown parameter θ. The maximum likelihood estimation method selects the set of values of the unknown parameter which maximize the joint probability density function of

the observations

$$\mathcal{L}(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

called the likelihood function.
Intuitively, these values maximize the "agreement" of the underling probability distribution with the observed data.

Theorem 1.3.2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having common density $f(\cdot\,;\theta)$. Assume that the support of $f(\cdot\,;\theta)$ is independent of θ and that $f(\cdot\,;\theta)$ is differentiable with respect to θ. Then the maximum likelihood estimator θ_n is a consistent estimator of θ, i.e. it converges almost surely (with probability 1) to θ [78]:*

$$\lim_{n \to +\infty} |\theta_n - \theta| = 0 \ \ a.s. \tag{1.46}$$

Here the subscript n denotes the sample size.

Non-parametric estimation

A commonly used non-parametric method to estimate the probability density function is the kernel density estimation.
Given a sample of n i.i.d. observations $(x_1, x_2, \ldots, x_n)$, the kernel density estimator of the underlying probability density function f is defined by

$$\hat{f}_n(x) := \frac{1}{n h_n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right), \tag{1.47}$$

where $h_n > 0$ is a smoothing parameter, depending on the sample size n, called bandwidth.
Here K is a symmetric non-negative function, called kernel function, such that

$$\int_{-\infty}^{+\infty} K(x)dx = 1,$$

$$\lim_{|x| \to +\infty} K(x) = 0.$$

In Table 1.2 we report some of the most common kernel functions.

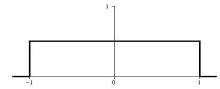
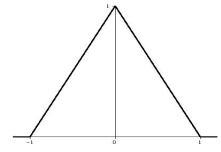
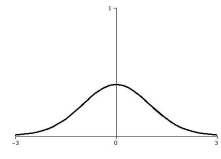
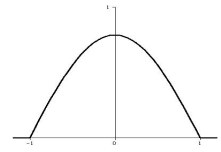
Uniform	$K(x) = \frac{1}{2}\mathbb{1}_{\{\lvert x\rvert \leq 1\}}$	
Triangular	$K(x) = (1 - \lvert x\rvert)\mathbb{1}_{\{\lvert x\rvert \leq 1\}}$	
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	
Cosine	$K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right) \mathbb{1}_{\{\lvert x\rvert \leq 1\}}$	

Table 1.2: The most common kernel functions. Here $\mathbb{1}$ denotes the indicator function.

Theorem 1.3.3. *Let us assume that the underlying probability density function f is continuous. Then taking a bounded variation kernel and choosing a suitable bandwidth sequence $\{h_n, n = 1, 2, \ldots\}$, such that*

$$\lim_{n \to +\infty} h_n = 0, \qquad \lim_{n \to +\infty} n h_n = +\infty$$

the kernel density estimator (1.47) is a uniform strongly consistent estimator of f, i.e. it converges uniformly and almost surely (with probability 1) to f:

$$\lim_{n \to +\infty} \sup_{x \in \mathbb{R}} \left| \hat{f}_n(x) - f(x) \right| = 0 \ \ a.s. \tag{1.48}$$

The bandwidth of a kernel density estimator is a free parameter which exhibits a strong influence on the resulting probability density function estimator. Indeed it

changes the rate of convergence of the estimator [116]. The most common optimality criterion used to select this parameter is to minimize the so called mean integrated squared error:

$$MISE = \mathbb{E}\left[\int_{-\infty}^{+\infty}\left(\hat{f}_n(x) - f(x)\right)^2 dx\right].$$

When Gaussian kernels are used and the underlying probability density function to estimate is Gaussian, then it can be shown that the optimal choice for h_n is

$$h_n = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}, \tag{1.49}$$

where $\hat{\sigma}$ is the sample standard deviation [116].

1.3.2 Estimators in presence of dependence

The probability density function estimators of Section 1.3.1 have good properties under weak conditions. Indeed these estimators converge to the unknown probability density function to estimate quickly as the sample size n tends to infinity (see Theorem 1.3.3 and 1.3.2, for instance). The same convergence properties still hold even if we relax the independence hypothesis. However in this case to ensure the almost sure convergence of the estimators, we require stronger condition on estimators and sample variables.
In this section we review some fundamental theorems that provide the uniform almost sure convergence of two non-parametric probability density function estimators, in presence of dependent sample variables.

Estimator for unconditional density functions

In [56], Gyorfi proved the almost sure L_2-convergence of a kernel-type probability density function estimator for ergodic processes. Many other researchers, [9, 60, 124] for instance, developed Gyorfi's work until in [23] the uniform almost sure convergence of a kernel-type probability density function estimator is proved. Here we report a specific version of this result adapted to our aims.

Let us consider an ergodic and stationary sequence $\{T_i, i = 1, 2, \ldots\}$ of inter-event intervals from a simple point process N, with shared probability density function f. A kernel-type estimator of this probability density function is defined as

$$\hat{f}_n(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{t - T_i}{b_n}\right), \ t \geq 0, \tag{1.50}$$

where the subscript n denotes the sample size.
Here $\{b_n\}$ is a sequence of positive real numbers such that

$$\lim_{n \to +\infty} b_n = 0, \qquad \lim_{n \to +\infty} nb_n = +\infty. \tag{1.51}$$

The function $K(t)$ is a kernel function on $\mathbb{R}$ such that

$$K(t) > 0 \ \forall t \in \mathbb{R}, \qquad \int_{\mathbb{R}} K(t)dt = 1, \qquad \lim_{|t| \to +\infty} K(t) = 0. \tag{1.52}$$

The uniform almost sure convergence of the kernel probability density function estimator (1.50) depends on the following assumptions. Here $f_i(t|\mathcal{F}_{i-1})$ is the conditional probability density function of T_i given the past history $\mathcal{F}_{i-1}$, defined in (1.34a).

A1: The densities $f(t)$ and $f_i(t|\mathcal{F}_{i-1})$ belong to the space $C_0(\mathbb{R})$ of real-valued continuous functions on $\mathbb{R}$ tending to zero at infinity.

A2: The conditional probability density functions $f_i(t|\mathcal{F}_{i-1})$, $i \geq 2$ are Lipschitz with ratio 1, i.e.

$$\left|f_i(t|\mathcal{F}_{i-1}) - f_i(t'|\mathcal{F}_{i-1})\right| \leq |t - t'| \quad \forall t, t' \in \mathbb{R} \ .$$

A3: The kernel $K(t)$ has bounded variation and it is Hölder with ratio $\mathfrak{L} < \infty$ and order $\gamma \in [0, 1]$:

$$\left|K(t) - K(t')\right| \leq \mathfrak{L}|t - t'|^{\gamma} \quad \forall t, t' \in \mathbb{R} \ .$$

Theorem 1.3.4. *Let $\{N(t), t \geq 0\}$ be a simple point process with ergodic and stationary inter-event intervals $\{T_i, i \geq 1\}$. Under assumptions **A1** to **A3**, for all sequences $\{b_n\}$ such that*

$$\lim_{n \to +\infty} \frac{nb_n^2}{\log n} = +\infty, \tag{1.53}$$

and any compact interval $[0, M] \subseteq \mathbb{R}_+$*, we have*

$$\lim_{n\to+\infty} \sup_{t\in[0,M]} \left|\hat{f}_n(t) - f(t)\right| = 0 \quad a.s. \tag{1.54}$$

Estimator for conditional density functions

Ould-Said [86] and Arfi [4] show the uniform almost sure convergence of a kernel-type estimator for conditional probability density functions. Here we provide a version of their result, adapted to our specific goals.

Let $\{T_i, i = 1, 2, \ldots\}$ be a stationary and ergodic sequence of inter-event intervals from a simple point process, with shared probability density function f. Here our aim is to provide a kernel-type estimator of the shared conditional density function $f(t|\tau)$ of an inter-event interval T_{i+1} given its preceding $T_i = \tau$. Let us recall that the formal definition of this probability density function is

$$f(t|\tau) = \frac{f(\tau, t)}{f(\tau)}, \tag{1.55}$$

where $f(\tau, t)$ is the shared joint probability density function of couples (T_i, T_{i+1}), $i \geq 1$, of subsequent inter-event intervals. Here the density function f plays the role of marginal probability density function of T_i, since the inter-event intervals are ergodic and stationary.
A kernel-type estimator of the conditional probability density function (1.55) is

$$\hat{f}_n(t|\tau) = \frac{\hat{f}_n(\tau, t)}{\hat{f}_n(\tau)}, \tag{1.56}$$

where

$$\hat{f}_n(\tau, t) = \frac{1}{nb_n^2} \sum_{i=1}^{n} K_1\left(\frac{\tau - T_i}{b_n}\right) K_2\left(\frac{t - T_{i+1}}{b_n}\right), \tag{1.57}$$

$$\hat{f}_n(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K_1\left(\frac{\tau - T_i}{b_n}\right). \tag{1.58}$$

Here the subscript n denotes the sample size and $\{b_n\}$ is a sequence of positive real numbers, such that $b_n \to 0$ and $nb_n \to +\infty$ as $n \to +\infty$. K_j, $j = 1, 2$, are kernel

functions such that $\lim_{|u|\to+\infty} K_j(u) = 0$. Moreover, we assume that these kernels have bounded variation and that K_1 is strictly positive.

The uniform almost sure convergence of the kernel probability density function estimator (1.56) depends on the following hypotheses. Here $f_i(\tau, t|\mathcal{G}_{i-1})$ is the conditional probability density function of the couple (T_i, T_{i+1}) given the joint past history $\mathcal{G}_{i-1}$, defined in (1.34b).

H1: The joint densities $f(\tau, t)$ and $f_i(\tau, t|\mathcal{G}_{i-1})$ belong to the space $C_0(\mathbb{R}^2)$ of real-valued continuous functions on $\mathbb{R} \times \mathbb{R}$ tending to zero at infinity.

H2: The marginal densities $f(t)$ and $f_i(t|\mathcal{F}_{i-1})$ belong to the space $C_0(\mathbb{R})$ of real-valued continuous functions on $\mathbb{R}$ tending to zero at infinity.

H3: The conditional density functions $f_i(\tau, t|\mathcal{G}_{i-1})$ and $f_i(t|\mathcal{F}_{i-1})$ are Lipschitz with ratio 1,

$$\left|f_i(\tau, t|\mathcal{G}_{i-1}) - f_i(\tau', t'|\mathcal{G}_{i-1})\right| \leq |\tau - \tau'| + |t - t'| \quad (\tau, \tau'),\ (t, t') \in \mathbb{R} \times \mathbb{R},$$

$$\left|f_i(t|\mathcal{F}_{i-1}) - f_i(t'|\mathcal{F}_{i-1})\right| \leq |t - t'| \quad (t, t') \in \mathbb{R} \times \mathbb{R}\ .$$

H4: Let $[0, M] \subseteq \mathbb{R}_+$ be a compact interval. We assume that for all t in an ϵ-neighbourhood $[0, M]_\epsilon$ of $[0, M]$ there exists $\gamma_\epsilon > 0$ such that $f(t) > \gamma_\epsilon$.

H5: The kernels K_j, $j = 1, 2$, are Hölder with ratio $\mathfrak{L} < \infty$ and order $\gamma \in [0, 1]$,

$$\left|K_1(\tau) - K_1(\tau')\right| \leq \mathfrak{L}|\tau - \tau'|^\gamma \quad (\tau, \tau') \in \mathbb{R} \times \mathbb{R}$$

$$\left|K_2(t) - K_2(t')\right| \leq \mathfrak{L}|t - t'|^\gamma \quad (t, t') \in \mathbb{R} \times \mathbb{R}$$

Remark 1.3.5. These assumptions are satisfied by any ergodic process with sufficiently smooth probability density functions (see [23] for details).

Theorem 1.3.6. *Let $\{N(t), t \geq 0\}$ be a simple point process with ergodic and stationary inter-event intervals $\{T_i, i \geq 1\}$. Under hypotheses **H1** to **H5**, for all sequence $\{b_n\}$ satisfying*

$$\lim_{n\to+\infty} \frac{nb_n^4}{\ln n} = +\infty\ , \tag{1.59}$$

and any compact interval $[0, M] \in \mathbb{R}_+$*, we have*

$$\lim_{n\to+\infty} \sup_{(\tau,t)\in[0,M]^2} \left|\hat{f}_n(t|\tau) - f(t|\tau)\right| = 0 \ \ a.s. \tag{1.60}$$

Remark 1.3.7. The bandwidth b_n strongly influences the rate of convergence of the kernel density estimators (1.50) and (1.56). However in case of dependent sample variables we don't know an optimality criterion to select these parameters yet, as in the classical case of i.i.d. sample variables.

CHAPTER 2

Neuronal background

The study of the nervous system dates back to the ancient Egypt. Manuscripts since 1700 B.C. indicate that the Egyptians had some knowledge about symptoms of brain damage. However the study of the brain has become a branch of science only after the invention of the microscope and the development of a staining procedure by Camillo Golgi during the late 1890s. This procedure uses a silver chromate salt to reveal the intricate structures of individual neurons.
The scientific study of the nervous system has had a new significant increase during the second half of the twentieth century, due to advances in electrophysiology and more recently in molecular biology. Immediately after the Second World War, Hodgkin and Huxley published the results of the first recording of an intracellular nervous signal, obtained by inserting a fine capillary electrode inside the nerve fibre of the giant axon of a squid. Nowadays neuroscientists are able to study the nervous system from different viewpoints.

For about a century the Golgi technique has been very successful in staining neurons. It enables the semi-automatic reconstruction and the quantitative analysis of their neuronal branching patterns (e.g. [53] and [87]). In recent times it is used in combination with other classical staining methods to achieve a quantitative statistical description of brain tissue in terms of the density of neurons, synapses, and total length of axons and dendrites.

Further progresses are supported by the spread of modern imaging techniques. They allow digital reconstructions of dendritic and axonal morphology. In the last few decades, the development of intracellular labelling [20, 120], using various visualization methods like fluorescent glyco-protein reaction [79], have led to a large output of high resolution data about dendritic morphology.
In addition, these new visualization approaches in conjunction with traditional in vitro (slice conditions) microscopy reduce the sampling bias due to limited access to some neuron types in vivo (anaesthetized animal) and variable cell survival during slice preparation.

However, in vivo visualizations enable the observation of developmental and activity-dependent morphological changes, such as individual spine plasticity [77]. Hence scientists are developing new visualization approaches also for the in vivo morphology observation, both at the whole cell and sub-cellular levels [85, 123].

In section 2.1 we revise some basic concepts of neuroanatomy and neurophysiology, useful for the understanding of the following chapters.

The human brain contains around 10^{10} neurons, connected to each other in complex networks. It is able to perform billions of activities, from motion to memory and learning. Part of the underlying processes to these neural functions are still unknown. The modern functional study of brain activities follows many directions. Some of the research topics of the last decades are, for instance:

- the functional brain mapping of epilepsy networks;

- the neural basis of social learning and social deciding;
- the neural mechanisms of the rewarding effects of drugs;
- the processes associated with fear memories.

All these research topics are based on the analysis of the relationship between external stimuli and neural responses.
Indeed neuroscientists already know that external stimuli are encoded by neurons in sequences of electrical pulses. However we still ignore completely the underlying coding mechanism for the information processing. The branch of neuroscience, which focuses on the mathematical problems in modelling the neural code of external stimuli, is computational neuroscience. There exists a wide literature on these topics (e.g. [47, 125, 126]), but many modelling features are still debated.
The aim of this thesis is to develop methods for the analysis of the dependency structures that arise during neural coding processes, from both the stochastic modelling (Chapters 3 and 5) and statistical (Chapter 6) point of view. In section 2.2 we introduce some neural models that are used in the following chapters.

2.1 Elements of neuroanatomy and neurophysiology

Neurons are the core components of the nervous system. They are highly specialized cells for the processing and transmission of information signals.
Given the wide range of different functions performed by humans and animals, there is a wide variety of neurons. However they share the same elementary structure (Figure 2.1), which is formed by three fundamental parts [125].

1. A focal part, called soma. It collects and elaborates input information from other neurons or the external environment.
2. A treelike structure, emanating from the soma, called dendritic tree. It is formed by the dendrites, which are the principal information gathering components of a neuron. Over the dendrites many contacts from other cells occur at specialized sites, called synapses.

3. A long projection of the nerve cell, called axon. It is responsible for the propagation of the information output signal from a cell to the other.

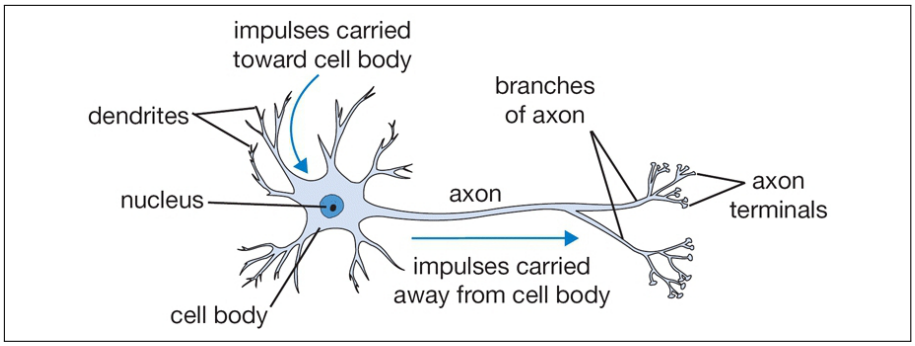


Figure 2.1: Schematic representation of the elementary structure of a neuron.
Source: http://www.wpclipart.com/medical/anatomy/cells/neuron/neuron.png.html

Neurons are electrically excitable cells, maintaining a voltage difference, called membrane potential, between the interior and the exterior of the cell. It is generated primarily from intracellular-versus-extracellular concentration differences of sodium, potassium, chloride and calcium ions, by means of metabolically driven ion pumps and channels, embedded in the neuron membrane.

In absence of neural activity, the membrane potential decays spontaneously to a characteristic level, called resting potential, which is about 70 mV negative inside. When an input arrives to a neuron from other cells or the external environment, it generates an alteration of ion concentrations, opening the ion channels. Whenever the membrane potential attains a critical level due to these alterations, an electrical pulse, called action potential, is elicited. Informally we say that the neuron fires and the membrane potential elicits a spike. After each spike, the membrane potential is reset to its resting value.

A chain of spikes emitted by a single neuron is called spike train (Figure 2.2), while the waiting time between two consecutive spikes is called interspike interval (ISI).

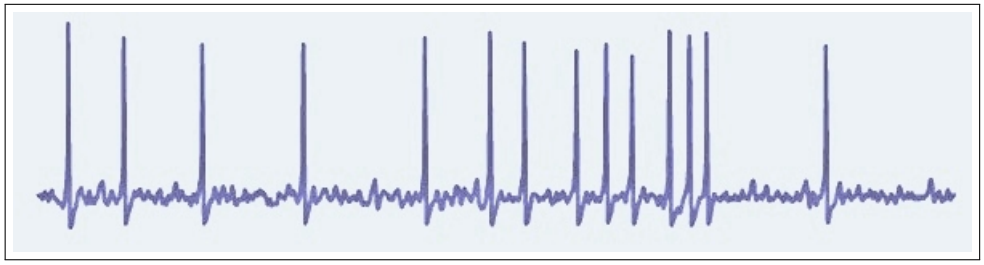


Figure 2.2: Example of spike train.
Source: http://www.oist.jp/press-room/news/2012/2/6/fiery-neurons

A traditional coding scheme, assumes that most of the information about external stimuli is contained in the firing rate of the neuron (e.g. [2, 3, 58, 80]). For instance, in motor neurons the strength at which an innervated muscle is flexed depends on the average number of spikes per time unit [46].

2.2 Neural modelling

The number of models for a single neuron dynamics is very large and their complexity ranges from oversimplified to highly realistic biophysical models.
The first mathematical model dates back to 1907, when Lapique ([71]) proposed to describe the membrane potential evolution of a neuron subject to an input, by using the time derivative of the physical law for the capacitance.
In the presence of an input current, the membrane potential increases until it reaches a constant threshold S. Then a spike occurs and the potential is reset to its resting value.

In 1952, Alan Lloyd Hodgkin and Andrew Huxley [57] introduced a physiologically detailed mathematical model for the transmission of electrical signals between neurons of the giant axon of a squid. This model treats each component of the nervous cell as an electrical element. The ion channels are represented by electrical conductances which depend on both voltage and time. The electrochemical gradients driving the flow of ions are represented by voltage sources, whose voltage is determined by

the ratio of the intra and extracellular ion concentrations. Finally, ion pumps are represented by current sources.

Although these models reasonably fit some experimental data, they are mathematically complex. The need of a model simplification leads to the birth of stochastic neural models. They separate the neural components in two groups: the principal components are accounted in a deterministic mathematical description of neuronal dynamics, while the others are globally summarized in a noise term.
The first attempt to formulate a stochastic neural model is due to Gerstein and Mandelbrot [45]. They describe the membrane potential dynamics before the release of an action potential through a Wiener process. Then, whenever the underlying stochastic process reaches a characteristic threshold, a spike occurs and the process is reset to its initial value.

This model is the basis of successive more realistic models. The most popular neural stochastic models are the one-dimensional Leaky Integrate and Fire (LIF) models (see [14], [15] and [111] for a depth review on these models). Their success is due to their relative simplicity jointly with their reasonable ability to reproduce neuronal input-output features.

2.2.1 LIF neural models

LIF models reproduce the membrane potential dynamics, between two consecutive spikes, through a one-dimensional stochastic process $X = \{X(t); t \geq 0\}$. It is characterized by a leakage term to model the spontaneous membrane potential decay in absence of inputs. Famous examples are the Stein's model and the Ornstein-Uhlenbeck diffusion model.

In [121], Stein formulates the first LIF model, where the membrane potential evolution is modelled by the process solution of the stochastic differential equation

$$\begin{cases} dX(t) = -\alpha\,(X(t) - \rho)\,dt + \delta^{+} dN^{+}(t) + \delta^{-} N^{-}(t) \\ X(t_0) = x_0 \end{cases} . \tag{2.1}$$

Here $\alpha > 0$ is the leakage constant, ρ is the resting membrane potential, $N^+(t)$ and $N^-(t)$ are independent Poisson processes of parameters λ^+ and λ^-, respectively, $\delta^+ > 0$ and $\delta^- < 0$ are the intensities of excitatory and inhibitory inputs.

The Ornstein-Uhlembeck diffusion model was proposed as a continuous limit of Stein's model [16]. Here the membrane potential evolution is modelled by the Ornstein-Uhlembeck process, defined by the stochastic differential equation

$$\begin{cases} dX(t) = (-\alpha X(t) + \mu)\, dt + \sigma dB(t) \\ X(t_0) = x_0 \end{cases} . \tag{2.2}$$

Here $\alpha > 0$ is the leakage constant, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ are the input intensity and variability, respectively and $B(t)$ is a standard Brownian motion.

The solution of (2.2) for a constant initial value x_0 is a one-dimensional Gauss-Markov diffusion process with mean

$$\mathbb{E}(X(t)) = \frac{\mu}{\alpha} + \left(x_0 - \frac{\mu}{\alpha}\right) e^{-\alpha t} \tag{2.3}$$

and variance

$$Var(X(t)) = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right). \tag{2.4}$$

Action potentials are elicited whenever the membrane potential process X exceeds, for the first time, a constant threshold S. After each spike, X is reset to its resting value $X(0) = x_0$ and the membrane potential evolution restarts according to the stochastic process dynamics (2.2).

Remark 2.2.1. For simplicity we suppose that the resting value x_0 of both the models is constant. However it is possible to find in the literature LIF models with time dependent resting membrane potential [13, 122].

Any ISI corresponds to the FPT

$$T_{S,x_0} = \inf\{t \geq 0 : X(t) \geq S \,|X(0) = x_0\} \tag{2.5}$$

of the stochastic process X across the boundary $S > x_0$. The assumed resetting mechanism ensures that ISIs are i.i.d. random variables, defining a so-called renewal

stochastic process. Therefore the knowledge of the ISI distribution corresponds to the knowledge of the distribution of the FPT T_{S,x_0}.
The FPT problem is a widely studied argument, which has a well-known solution in some simple cases, but in general it is still an open problem, subject of many theoretical studies [12, 24, 93, 115].
The analytical solution is known in very few instances, like the Weiner process of Gerstein and Mandelbrot model with constant threshold [32]. On the contrary for the Ornstein-Uhlenbeck process we only know the Laplace transform of its FPT probability density function. The inverse transform can be derived in closed form in very few instances [113], but there exist some efficient numerical methods to approximate the FPT probability distribution and its moments for an Ornstein-Uhlenbeck process [88, 98].

LIF model popularity is due to its mathematical tractability [29]. It derives from the fact that they concentrate the neuron features into a single point. This implies to disregard completely the geometrical structure of a neuron, losing some important features of the neuron dynamics. For instance, LIF models are not able to reproduce the experimentally observed dependence between successive ISIs.
Attempts to generalize one-dimensional LIF models make use of the LIF paradigm in the frame of multi-compartment models. In [10], [65], [67] and [68], two-compartment neural models are discussed.

2.2.2 Two-compartment neural models

Two-compartment neural models are spatially complex models. They model the dynamics of neural signal flows between two interconnected parts of a neuron. They are described by two-dimensional diffusion processes with linear drift,

$$d\mathbf{X}(t) = \{A\,\mathbf{X}(t) + \mathbf{M}(\mathbf{t})\}\,dt + \Sigma(t)\,d\mathbf{B}(t), \tag{2.6}$$

were A and $\Sigma(t)$ are 2×2-matrices, $\mathbf{M}(t)$ is a two-dimensional vector and $\{\mathbf{B}(t), t \geq 0\}$ is a two-dimensional standard Brownian motion.
The components of the two-dimensional diffusion process $\{\mathbf{X}(t), t \geq 0\}$ model the

membrane potentials in each compartment. $\mathbf{M}$ and Σ represent the intensity and the variability of external inputs to the neuron, respectively.

In this thesis we consider the two-compartment neural model proposed in [69]. Here the two compartments correspond to the dendritic tree and the soma. The dendritic component is responsible for receiving external inputs, while the somatic component emits outputs. Hence we assume that external inputs reach indirectly the soma by the interconnection between the two compartments (Figure 2.3).

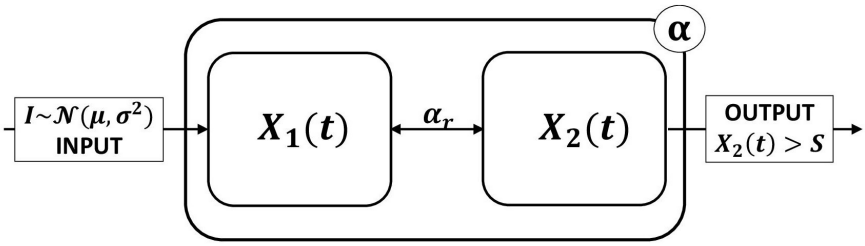


Figure 2.3: A schematic representation of a two-compartment neural model.

The model is described by a two-dimensional diffusion process $\{\mathbf{X}(t), t \geq 0\}$, whose components $X_1(t)$ and $X_2(t)$ model the membrane potential evolution in the dendritic and somatic compartment, respectively. Assuming that external inputs have intensity μ and variance σ^2, we obtain the following two-compartment model

$$dX_1(t) = \{-\alpha X_1(t) + \alpha_r [X_2(t) - X_1(t)] + \mu\}dt + \sigma dB(t) \tag{2.7a}$$

$$dX_2(t) = \{-\alpha X_2(t) + \alpha_r [X_1(t) - X_2(t)]\}dt \tag{2.7b}$$

with $X_1(0) = x_1$ and $X_2(0) = x_2$. Here α_r is the strength of the interconnection between the two compartments, while α is the leakage constant that models the spontaneous membrane potential decay in absence of inputs. For the sake of simplicity, we assume that the membrane time constants are the same in both the compartments, however this assumption can be easily removed.

The neuron dynamics is modelled in the following way. Whenever the somatic membrane potential X_2 reaches a characteristic threshold S the neuron elicits a spike. Then the value of X_2 is reset to its resting value while the dendritic membrane potential X_1 continues its evolution.

In absence of a firing threshold, the solution of (2.7a) and (2.7b) for constant initial values is a bivariate Gauss-Markov diffusion process with mean $\mathbf{m}(t) = \mathbb{E}(\mathbf{X}(t))$, whose components are:

$$m_1(t) = m_1(\infty) + \frac{1}{2}\left(x_1 + x_2 - \frac{\mu}{\alpha}\right) a(t) + \frac{1}{2}\left(x_1 - x_2 - \frac{\mu}{\alpha + 2\alpha_r}\right) b(t), \tag{2.8a}$$

$$m_2(t) = m_2(\infty) + \frac{1}{2}\left(x_1 + x_2 - \frac{\mu}{\alpha}\right) a(t) + \frac{1}{2}\left(x_2 - x_1 + \frac{\mu}{\alpha + 2\alpha_r}\right) b(t), \tag{2.8b}$$

where $a(t) = e^{-\alpha t}$ and $b(t) = e^{-(\alpha+2\alpha_r)t}$.

The initial membrane potentials x_1 and x_2 are identified with the resting potentials of both the compartments, when the time origin coincides with a firing epoch and the first component is in a stationary regime. For notational simplicity we identify the resting potential with zero. It is only a translation of the model values, which does not change the results.

The constants

$$m_1(\infty) = \frac{(\alpha + \alpha_r)\mu}{\alpha(\alpha + 2\alpha_r)} \quad , \quad m_2(\infty) = \frac{\alpha_r \mu}{\alpha(\alpha + 2\alpha_r)} \tag{2.9}$$

represent the asymptotic mean membrane potentials. Note that the membrane potential of the dendritic zone is always greater than the one of the somatic compartment. Furthermore the membrane potentials of the two components become similar when $\alpha_r \gg \alpha$.

When the initial values are constant, the covariance matrix $\Gamma(t, \tau)$ of the bivariate process $\mathbf{X}$ has components

$$\Gamma_{11}(t,\tau) = Var\left(X_1(\infty)\right) - \frac{2\alpha_r^2 c(t-\tau)^2 + \alpha\alpha_r\left(1+4c(t-\tau)+3c(t-\tau)^2\right) + \alpha^2\left(1+c(t-\tau)\right)^2}{8\alpha(\alpha+\alpha_r)(\alpha+2\alpha_r)d(t-\tau)}\sigma^2\,, \tag{2.10}$$

$$\Gamma_{22}(t,\tau) = \Gamma_{11}(t,\tau) - \frac{\sigma^2}{2(\alpha+\alpha_r)} + \frac{\sigma^2}{2(\alpha+\alpha_r)d(t-\tau)}\,,$$

$$\Gamma_{12}(t,\tau) = Cov(X_{12}(\infty)) - \frac{(\alpha+2\alpha_r)c(t-\tau)^2 - \alpha}{8\alpha(\alpha+2\alpha_r)d(t-\tau)}\sigma^2\,, \tag{2.11}$$

$$\Gamma_{21}(t,\tau) = \Gamma_{12}(t,\tau)\,, \tag{2.12}$$

where $c(t) = e^{2\alpha_r t}$ and $d(t) = e^{2(\alpha+\alpha_r)t}$.

Here the constants

$$Var\left(X_1(\infty)\right) = \frac{(2\alpha^2 + 4\alpha\alpha_r + \alpha_r^2)\sigma^2}{4\alpha(\alpha+\alpha_r)(\alpha+2\alpha_r)} \tag{2.13}$$

and

$$Cov(X_{12}(\infty)) = \frac{\alpha_r\sigma^2}{4\alpha(\alpha+2\alpha_r)} \tag{2.14}$$

denote the asymptotic dendritic variance and covariance between the two compartments, respectively.

Remark 2.2.2. Besides their use to model neurons, two-compartmental models are applied in many fields, including pharmacokinetics, epidemiology, biomedicine, systems theory, complexity theory, engineering, physics, information science and social science [54, 75, 81, 112], to model the flow of substances between two interconnected parts of a system.

CHAPTER 3

Dependency structure of a single spike train generated by a two-compartment neural model.

The two-compartment model, introduced in Section 2.2.2 has something in common with the Ornstein-Uhlenbeck one-dimensional LIF model described in [96]. However, the lack of resetting of the dendritic component destroys the renewal character of one-dimensional models (Figure 3.1). Hence dependent ISIs are generated by this model, due to the absence of the renewal hypothesis. For particular choices of the model parameters there is a statistical evidence of this dependence, as we highlight in [7].

In this Chapter we perform an analysis of the new neural features reproduced by the

two-compartment neural model from a statical point of view.
In many cases we require the statistical stationarity of the dendritic component. Indeed we prove the identical distribution of the dendritic component at different spiking times, through suitable statistical tests.

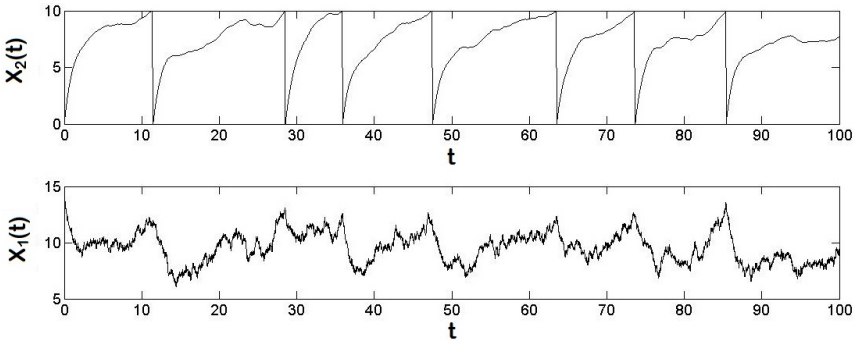


Figure 3.1: Example of evolution of the dendritic (lower panel) and somatic (upper panel) membrane potentials, simulated according with equations (2.7a) and (2.7b). Here $\alpha = 0.05\ ms^{-1}$, $\alpha_r = 0.5\ ms^{-1}$, $\mu = 1.5\ mV$, $\sigma = 1\ mV/ms^{1/2}$, and $S = 10\ mV$. Note that the dendritic component is not reset after a spike.

In the following, we denote with i^* the index of the spiking epoch at which the dendritic component $X_1(t)$ is statistically stationary. To estimate the value of i^* we perform Kolmogorov-Smirnov tests ([78], Ch. 11) on the random variables $X_1(l_i)$ and $X_1(l_{i+1})$, $i \geq i^*$, to check their identical distribution.

3.1 Model dynamics

We assume that our origin of times coincides with the epoch of a spike. We indicate with l_i the epoch of the i-th spike, $i \geq 1$. Then the i-th ISI, $i \geq 1$, is described by the random variable

$$T_i = \inf \{t > 0 : X_2(t) \geq S \,|X_2(l_{i-1}) = 0\}, \tag{3.1}$$

with $l_0 = 0$ and $T_1 = l_1$.

To study the dynamics of this model we separate the case of absence of noise from the one with noise, following a classical approach of one-dimensional models.

3.1.1 Absence of Noise

When $\sigma = 0$, the time evolution of the dendritic and somatic potentials is given by equations (2.8a) and (2.8b), respectively. In the subthreshold regime ($S > m_2(\infty)$) the neuron is silent, while in the suprathreshold regime ($S < m_2(\infty)$) it spikes regularly at fixed times $l_j = l_{i^*} + (j - i^*)T_{i^*}$, for $j \geq i^*$.
If the spike frequency is low, the two components attain their stationary dynamics during each ISI. Then we have $m_1(l_i) = m_1(\infty)$, $m_2(l_i) = S$ and $m_2\left(l_i^+\right) = 0$, for $i \geq 1$. Here l_i^+ indicates the instant immediately following the i-th spike.
In the case of supra-threshold regime and low spiking frequency, with initial condition $x_1 = m_1(\infty)$ and $x_2 = 0$, each ISI T_i, $i \geq 1$, is solution of the equation (see [69])

$$S - m_2(\infty) = \left(e^{-\alpha T_i} + e^{-(\alpha+2\alpha_r)T_i}\right) \frac{S - m_2(\infty)}{2} + e^{-2(\alpha+\alpha_r)T_i} m_2(\infty). \qquad (3.2)$$

Equation 3.2 relates the ISIs with the asymptotic somatic membrane potential, whenever the dendritic component attains its stationary dynamics during each ISI.
Since the dendritic potential evolution is perturbed by the resetting of the somatic component, the stationary regime is not attained during the first ISI. Therefore, in general (3.2) holds for any T_j, with $j > i^*$.

3.1.2 Presence of Noise

For $\sigma > 0$, the value of the dendritic component at spiking epochs is random and its distribution depends upon the preceding dynamics of the process. Hence a dependency between ISIs and the past evolution of the membrane potential appears.
When the dendritic component is stationary, approximate formulas relating ISIs and the values of the dendritic component at spiking epochs can be proved. To obtain these formulas we integrate equation (2.7b) between two spiking epochs, l_{i-1} and l_i. Note that the somatic component can not attain values larger than S on $t \in (l_{i-1}, l_i)$,

$i \geq 1$. Hence we introduce the process $X_2^B(t)$, with $t \in (l_{i-1}, l_i)$ to model the somatic membrane potential. The sample paths of $X_2^B(t)$ coincide with those of $X_2(t)$ that have not crossed S on $t \in (l_{i-1}, l_i)$, $i \geq 1$. Then, by definition, $X_2^B(l_{i-1}^+) = 0$ and $X_2^B(l_i) = S$. Formally, $X_2^B(t)$ is a Bridge process not crossing the boundary S on $t \in (l_{i-1}, l_i)$, described in [51]. For $t \in (l_{i-1}, l_i)$ it is solution of

$$X_2^B(l_i) - X_2^B(l_{i-1}) = -(\alpha + \alpha_r) \int_{l_{i-1}}^{l_i} X_2^B(t)dt + \alpha_r \int_{l_{i-1}}^{l_i} X_1(t)dt\,. \tag{3.3}$$

In order to determine a relationship between the value of the dendritic component at l_{i-1} and the ISI T_i,$i \geq 1$, we separate the analysis of (3.3) into the two cases of sub-threshold and supra-threshold regimes.

Supra-threshold regime

When $m_2(\infty) > S$, i.e. when input are strong, ISIs are short and $X_2^B(t)$ can be approximated by $X_2(t)$ for $t \in (l_{i-1}, l_i)$, with $X_2(l_i) = S$. Indeed, in this case, multiple crossings of the threshold on short time intervals are rare and a small percentage of sample paths of $X_2(t)$ has not a corresponding sample path of $X_2^B(t)$. Hence equation (3.3) can be rewritten as

$$X_2(l_i) - X_2(l_{i-1}) = -(\alpha + \alpha_r) \int_{l_{i-1}}^{l_i} X_2(t)dt + \alpha_r \int_{l_{i-1}}^{l_i} X_1(t)dt\,. \tag{3.4}$$

Taking the expectation of each member of (3.4) and applying Fubini's theorem [40], we get

$$S = -(\alpha + \alpha_r) \int_{l_{i-1}}^{l_i} m_2(t)dt + \alpha_r \int_{l_{i-1}}^{l_i} m_1(t)dt. \tag{3.5}$$

When the spiking activity is fast, the dendritic component does not attain its stationary regime during each ISI and it assumes different values at spiking epochs. Hence the value of the dendritic membrane potential at time l_i depends upon the past dynamics of the process, $\{\mathbf{X}(t), t < l_i\}$.

We denote with $M_i = \mathbb{E}\left[X_1\left(l_i\right) | \mathbf{X}(t), t < l_i\right]$, the expected value of the dendritic component at a spiking epoch, conditioned upon the previous history of the process. Then the expressions of (2.8a) and (2.8b), with initial conditions $x_2 = m_2(l_{i-1}) = 0$

and $x_1 = m_1(l_{i-1}) = M_{i-1}$, are

$$m_1(t) = m_1(\infty) + \frac{1}{2}\left(M_{i-1} - \frac{\mu}{\alpha}\right)e^{-\alpha t} + \frac{1}{2}\left(M_{i-1} - \frac{\mu}{\alpha + 2\alpha_r}\right)e^{-(\alpha+2\alpha_r)t}, \quad (3.6a)$$

$$m_2(t) = m_2(\infty) + \frac{1}{2}\left(M_{i-1} - \frac{\mu}{\alpha}\right)e^{-\alpha t} + \frac{1}{2}\left(\frac{\mu}{\alpha + 2\alpha_r} - M_{i-1}\right)e^{-(\alpha+2\alpha_r)t}. \quad (3.6b)$$

Replacing (3.6a) and (3.6b) into (3.5), we get the following approximate formula relating ISIs and values of the dendritic component at spiking epochs:

$$2(S - m_2(\infty)) = \left(M_{i-1} - \frac{\mu}{\alpha}\right)e^{-\alpha T_i} + \left(\frac{\mu}{\alpha + 2\alpha_r} - M_{i-1}\right)e^{-(\alpha+2\alpha_r)T_i}. \quad (3.7)$$

When $\alpha \to 0$, $e^{-\alpha T_i} \approx 1$ and $1 - \frac{2S}{M_{i-1} - \frac{\mu}{\alpha+2\alpha_r}} > 0$, equation (3.7) can be solved to get

$$T_i \approx -\frac{1}{\alpha + 2\alpha_r} \ln\left(1 - \frac{2S}{M_{i-1} - \frac{\mu}{\alpha+2\alpha_r}}\right) \quad i = 1, 2, \ldots \quad (3.8)$$

Hence the distribution of the $i - th$ ISI T_i depends on the entire past history of the process, through the conditional expectation M_{i-1} of the dendritic component at the previous spiking epochs.

However, when the conditional random variables M_j, $j \geq 1$, are identically distributed and their distribution does not depend upon the previous evolution of the process, the ISIs T_{j+1}, $j \geq 1$, become identically distributed. Indeed this happens for any $j \geq i^*$, when the dendritic component is stationary. In this case the ISI distribution coincides with the distribution of the FPT of the somatic component through the threshold S.

Moreover, the ISIs $T_i, T_{i+1}, ..., T_{i+n}$, $i > 1$ are dependent since M_{j-1} depends upon T_{j-1}, $j = i, \ldots, i + n$.

A further approximation of (3.8),

$$T_i \approx \frac{1}{\alpha + 2\alpha_r}\left(\frac{2S}{M_{i-1} - \frac{\mu}{\alpha+2\alpha_r}}\right), \quad (3.9)$$

holds when $\frac{2S}{\frac{\mu}{\alpha+2\alpha_r}-M_{i-1}}$ is small enough, i.e. for large inputs. Hence the mean firing frequency is approximately

$$\mathbb{E}\left(\frac{1}{T_i}\right) \approx \frac{\alpha+2\alpha_r}{2S}\left(\mathbb{E}(M_{i-1}) - \frac{\mu}{\alpha+2\alpha_r}\right) \tag{3.10}$$

and its variance is

$$Var\left(\frac{1}{T_i}\right) \approx \left(\frac{\alpha+2\alpha_r}{2S}\right)^2 Var(M_{i-1}). \tag{3.11}$$

Furthermore for the correlation we get:

$$\rho\left(\frac{1}{T_i}, \frac{1}{T_{i-1}}\right) \approx \rho(M_{i-1}, M_{i-2}). \tag{3.12}$$

Formulas (3.8)-(3.12) are not useful for computational aims, as their use requests the knowledge of the moments of the conditional random variable M_i, $i \geq 1$. However they are interesting because they illustrate the relationship between the moments of the random variables T_i and M_i.

Sub-threshold regime

When the somatic membrane potential is in the sub-threshold regime, formulas from (3.7) to (3.12) do not hold. However, in this case, the attainment of the threshold is rare and it is determined by the noise. With moderate noise intensity, ISIs increase and the dendritic component attains its stationary regime during each ISI. Hence we can postulate the identical distribution of T_i, $i \geq 1$. Furthermore in this case, during each ISI, the process forgets the initial value of the dendritic component. Hence the ISIs are approximately i.i.d. The presence of the renewal property makes the features of the two-compartment model similar to those of one-dimensional LIF models. Hence our interest focuses mainly on the supra-threshold regime.

3.2 Model features

To discuss the dependency features of the model we make use of the approximated formulas of the previous section and of simulations. Here we focus on the dependency

properties between successive ISIs, as the parameter values vary, while we refer to [69] for further properties.

We first perform a sensitivity analysis on the parameters α_r, σ and μ involved in the model. We recognize that particular choices of these parameters make the ISIs dependent but identically distributed. Then we compare the proposed two-compartment model with the more popular one-dimensional LIF models.
To quantitatively compare these models, we should select the criteria to fix the parameter values. A reasonable choice is to estimate the parameter values from recorded data for each model. Unfortunately this is not a simple task.

Recent papers, like [26–28, 70], deal with the parameter estimation problem for one-dimensional LIF models. These results can be generalized to estimate μ, σ and α. Indeed μ and σ are related to the input mean and variability, while α can be identified, as in the case of the Ornstein-Uhlenbeck model [96], with the transmembrane leakage. However, the estimation problem for α_r is new. It is an abstract meaning parameter that models the connection between the compartments. Hence, suitable statistical methods should be developed to estimate this parameter. However, this task requires further mathematical efforts, postponed to future works. Hence the comparison between the proposed model and one-dimensional LIF models is just qualitative.

The discussion of joint and marginal ISI distributions is postponed to Chapter 5.

3.2.1 Role of the parameters

In this section we provide a comprehensive description of the new neural features modelled by the two-compartment model. The aim of this analysis is to show the different roles of the model parameters on the neural dynamics. Hence the selection of the model parameter values is not suggested by any attempt to reproduce realistic instances.

Where not differently established, the parameters values are: $S = 10\ mV$, $\alpha = 0.05$ ms^{-1}, $\alpha_r = 0.5\ ms^{-1}$, $\sigma = 1\ mV/ms^{1/2}$, $\mu = 3.5\ mV$. We use simulations of 1000 sample paths.

Role of α_r

The junctional constant α_r determines the strength of the connections between the two compartments. When $\alpha_r = 0$, the somatic potential evolves independently from the dendritic one. Actually its dynamics becomes deterministic, because it does not receive noise from the dendritic component.

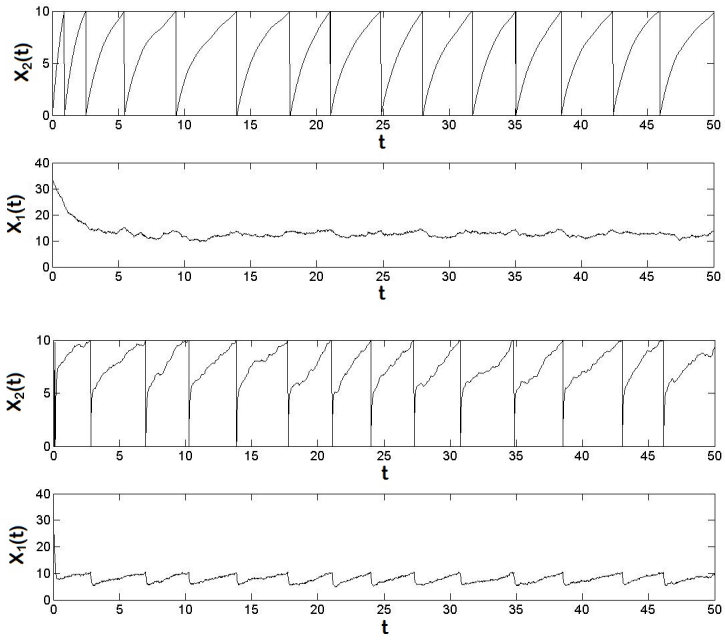
For fixed α, as α_r increases, the dependence between the values of the dendritic component at successive spiking epochs decreases. Indeed, when α_r increases, the somatic potential dynamics strongly affects the dendritic potential evolution. Hence, for larger values of α_r, both potentials exhibit a resetting effect at spiking epochs and the ISIs become i.i.d (see Figure 3.2(b)). On the contrary, in the presence of a weak coupling between the two compartments, the dendritic potential can attain a stationary dynamics, as little perturbation arrives from the somatic dynamics. Hence the renewal feature affects only the somatic component, generating a dependence between successive ISIs.

α_r	$\hat{\tau}$	$\hat{\rho}$	i^*
0.05	$[0.39, 0.47]$	$[0.57, 0.65]$	4
0.25	$[0.11, 0.19]$	$[0.17, 0.29]$	5
0.5	$[0.10, 0.18]$	$[0.10, 0.22]$	5
0.75	$[0.03, 0.11]$	$[0.05, 0.16]$	5

Table 3.1: Estimated values of Kendall's τ and correlation coefficient ρ between subsequent ISIs (95% confidence intervals).

In Table 3.1 we illustrate the dependence between two successive ISIs reporting the values of the Kendall's τ and the correlation coefficient ρ. The estimated values $\hat{\tau}$

and $\hat{\rho}$ refer to successive ISIs. With the choice of the parameters of Table 3.1, the ISI T_{i^*} and T_{i^*+j}, $j > 1$, are dependent ($\hat{\tau} > 0.1$) when $\alpha_r = 0.05$, otherwise the dependence disappears (see Table 3.2).



(a)

(b)

Figure 3.2: Two examples of samples of the somatic (upper) and dendritic (lower) components. In panels (a) $\alpha_r = 0.5\ ms^{-1}$, while in panels (b) $\alpha_r = 10\ ms^{-1}$.

	(T_i^*, T_{i^*+1})	(T_i^*, T_{i^*+2})	(T_i^*, T_{i^*+3})	(T_i^*, T_{i^*+4})
$\hat{\tau}$	$[0.57, 0.64]$	$[0.23, 0.34]$	$[0.10, 0.22]$	$[-0.04, 0.08]$
$\hat{\rho}$	$[0.39, 0.46]$	$[0.14, 0.22]$	$[0.06, 0.14]$	$[-0.03, 0.05]$

Table 3.2: Estimated values of Kendall's τ and correlation coefficient ρ (95% confidence intervals) between different couples of ISIs. Here $\alpha_r = 0.05$ and $i^* = 4$. Note that for $\alpha_r = 0.05$ T_{i^*} and T_{i^*+j} are dependent ($\hat{\tau} > 0.1$) for $j \in [1, 3]$.

Role of σ

The noise affects directly only the dendritic compartment. However the interconnection between the two compartments allows the input variability to influence the somatic dynamics and the distribution of ISIs. Indeed, increasing σ, the ISI variability increases, as shown in Table 3.3.

σ	i^*	$\sigma^2(T_6)$	$\hat{\tau}$	$\hat{\rho}$
0.05	6	$[0.003, 0.004]$	$[0.06, 0.14]$	$[0.07, 0.19]$
1	5	$[0.33, 0.39]$	$[0.08, 0.16]$	$[0.13, 0.25]$
5	4	$[9.21, 10.98]$	$[0.09, 0.17]$	$[0, 0.11]$
10	3	$[26.93, 32.09]$	$[0.21, 0.29]$	$[0.03, 0.15]$

Table 3.3: Estimated values of i^*, Kendall's τ (95% confidence intervals), correlation coefficient ρ (95% confidence intervals) between successive ISIs and sample variance $\sigma^2(T_6)$ of T_6 (95% confidence interval), for different noise intensities. i^* is determined using a Kolmogorov-Smirnov test ([78], Ch. 11) to check the identical distribution of $X_1(l_i)$ and $X_1(l_{i+1})$ for $i \geq i^*$.

Furthermore, for increasing values of σ, the stationary distribution of M_i, $i \geq i^*$ becomes flatter (Figure 3.3).
As a consequence, the dependence between successive ISIs increases (see Table 3.3). However the dependence disappears when one considers T_{i^*} and T_{i^*+j}, $j > 1$.

From Table 3.3 we can also observe that the Kendall's τ captures the ISI dependencies better than correlation coefficient ρ. This fact is related to the properties of the two dependence measures: ρ detects linear dependencies while τ does not hypothesize specific shapes of the dependencies. Furthermore ρ is the ratio of a covariance with the product of two standard deviations. Both the ISI covariance and standard deviations increase with σ. This causes the non monotonic behaviour of ρ.

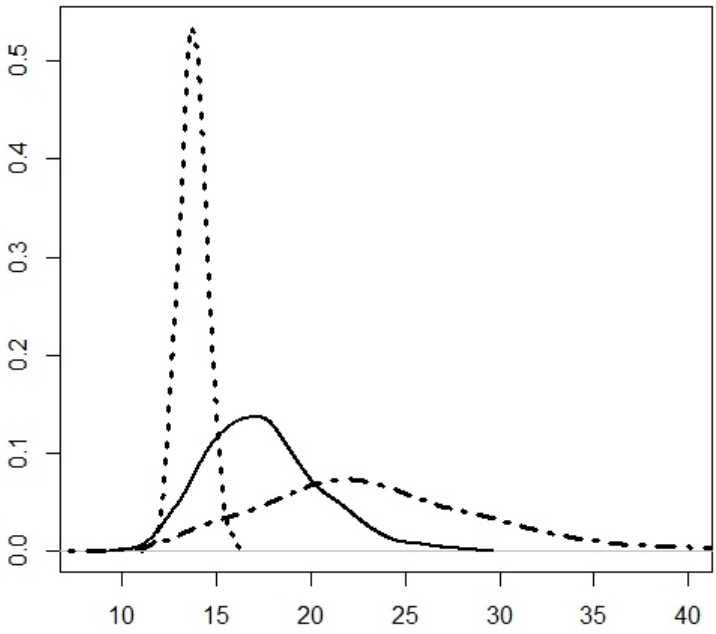


Figure 3.3: Stationary probability density function of M_i, $i \geq i^*$, for different values of σ: $\sigma = 1$ (dotted), $\sigma = 5$ (solid), $\sigma = 10$ (dashdot). Note that for increasing values of σ, the distribution becomes flatter.

Role of μ

Examples in Table 3.4 show that the dependence between successive ISIs increases with the mean input μ.

For small values of the input μ the somatic component is in the sub-threshold regime. In this case the neuron is slow and the somatic component attains its stationary regime during each ISI. Furthermore during each ISI, the process has the necessary time to forget the initial value of the dendritic component and ISIs are independent (Table 3.4, third and fourth column). For values of $\mu < 2\ mV$, a Kolmogorov-Smirnov test ([78], Ch. 11) on the distribution of M_i, $i \geq 1$, confirms that these random variables are identically distributed (with a p-value of 0.29).

μ	i^*	$\hat{\tau}$	$\hat{\rho}$	$\mu(T_j)_{j>i^*}$	$m_2(\infty) - S$
1	1	$[-0.05, 0.03]$	$[-0.05, 0.07]$	52.401 ms	-0.48
2	2	$[-0.02, 0.06]$	$[-0.05, 0.07]$	8.7091 ms	9.05
3	4	$[0.06, 0.14]$	$[0.10, 0.22]$	4.7324 ms	18.57
4	6	$[0.16, 0.24]$	$[0.20, 0.32]$	3.2923 ms	28.09
5	8	$[0.34, 0.42]$	$[0.33, 0.44]$	2.5176 ms	37.62

Table 3.4: Estimated values of i^*, Kendall's τ (95% confidence intervals) and correlation coefficient ρ (95% confidence intervals) and ISI sample mean $\mu(T_j)$ for $j > i^*$ as μ varies. In the last column the values of $m_2(\infty) - S$ allow to recognize sub-threshold and supra-threshold regimes.

When the input μ increases, the ISIs decrease and $X_1(t)$ does not attain its stationary regime during the first ISI. This implies that the variables M_i, $i \geq 1$, are not identically distributed for small values of i. However for $i \geq i^*$ the random variables M_i become identically distributed (with a p-value of 0.53). Furthermore as μ increases, successive ISIs, as well as successive values of the variables M_i, become dependent. This dependence strengthens with μ, as shown by Table 3.4. This fact can be explained considering the decrease of the ISIs as μ increases. Indeed, the process does not forget its starting point when the spikes are frequent. In particular for $\mu > 4$ we can also observe a light dependence between ISIs T_{i^*} and T_{i^*+j}, $j > 1$ (see Table 3.5).

μ	i^*	(T^*_i, T_{i^*+1})	(T^*_i, T_{i^*+2})	(T^*_i, T_{i^*+3})
5	8	$[0.33, 0.44]$	$[0.05, 0.17]$	$[-0.05, 0.07]$
7.5	12	$[0.51, 0.60]$	$[0.16, 0.22]$	$[0.04, 0.16]$
10	17	$[0.62, 0.69]$	$[0.27, 0.38]$	$[0.10, 0.22]$

Table 3.5: Estimated values of Kendall's τ (95% confidence intervals) between different couples of ISIs with $\mu > 4$

3.2.2 Qualitative comparison with LIF models

The ability to reproduce many qualitative and quantitative features of data, combined with their relative simplicity, has determined the popularity of LIF models. However their strong simplification of the neural structure has two main deficiencies:

- the geometry of the neuron is not considered;
- the ISIs are independent and identically distributed random variables.

The proposed two-compartment neural model is a first attempt to investigate the effects of the interaction between different parts of the neuron. It is characterized by several features that distinguish it from classical LIF models.

The indirect transfer of the input signal from the dendritic to the somatic component causes a delayed reaction of the modelled neuron, as it is apparent from Figure 3.1. Then, the two-compartment neural model is more stable with respect to negligible short changes in the input intensity, as the effects of noise on the somatic compartment are filtered by the connection between the compartments. Indeed the variance of the somatic compartment is always smaller then the variance of the dendritic compartment, as well illustrated by the relationship between the asymptotic variances of the two compartments [69]:

$$Var(X_2(\infty)) = Var(X_1(\infty)) - \frac{\sigma^2}{2(\alpha + \alpha_r)} < Var(X_1(\infty)). \tag{3.13}$$

Here $Var(X_2(\infty))$ is the asymptotic somatic variance and $Var(X_1(\infty))$ is the asymptotic dendritic variance, defined in (2.13).
Moreover the two-compartment neural model responds to input variations by a lower variability in the ISI distribution (see Table 3.3, first and third columns). Indeed the asymptotic variances $Var(X_1(\infty))$ and $Var(X_2(\infty))$ are always smaller then the asymptotic variance of a LIF model. For instance we have

$$\frac{Var(X(\infty))}{Var(X_2(\infty))} = 4 + 6\frac{\alpha}{\alpha_r} + 2\left(\frac{\alpha}{\alpha_r}\right)^2 > 4 \tag{3.14}$$

where $Var(X(\infty))$ is the asymptotic variance of a LIF model, obtained taking the limit of (2.4) as t tends to $+\infty$.

A further important feature is the serial dependence between ISIs, not observed in spike trains generated by any one-dimensional LIF model. Indeed the renewal hypothesis of LIF models prevents any relationship between successive ISIs.
The serial dependence between ISIs characterizes bursting and clustering activity (Figure 3.4), often experimentally observed [17, 22, 106]. The role of bursting activity for information processing is apparently enormous and the two-compartment model is a simple example of neural model achieving bursting for constant inputs. Moreover in Section 3.2.1 we show that the dependence between ISIs increases when μ increases. Hence this model suggests a possible mechanism which determines bursts and clusters: a sufficiently strong input.

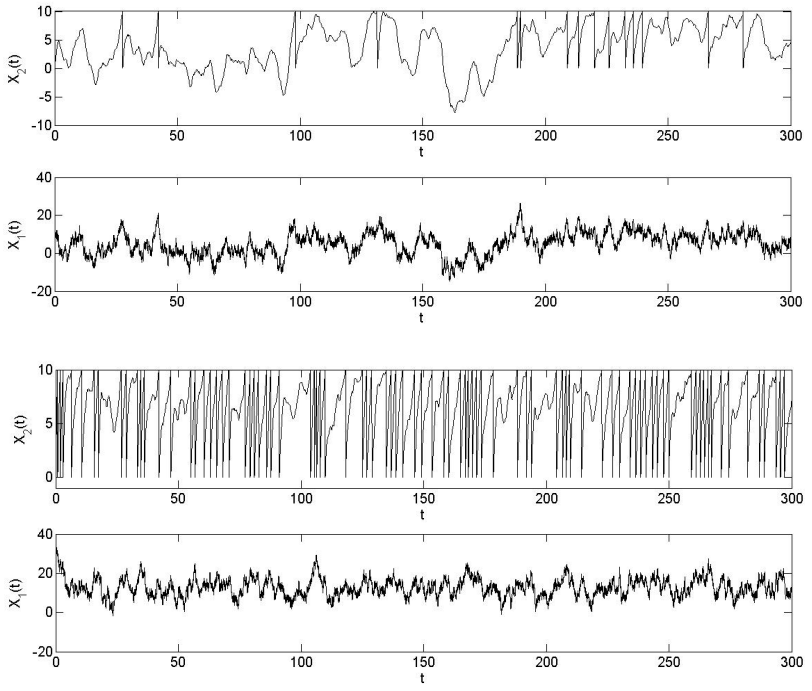


Figure 3.4: Examples of evolution of the two neural compartments for different values of μ: $\mu = 1\ mV$ (a) and $\mu = 3.5\ mV$ (b). Note that figure (b) shows bursting activity due to the increase of the input intensity. The other parameters are $\alpha = 0.05\ ms^{-1}$, $\alpha_r = 0.5\ ms^{-1}$, $\sigma = 5\ mV/ms^{1/2}$ and $S = 10\ mV$.

CHAPTER 4

A FPT problem for bivariate stochastic processes: a numerical solution

FPT problems arise in a variety of applications ranging from finance to biology, physics or psychology ([93, 95, 119] and examples cited therein).
Analytical [48, 49, 88, 91, 97, 109], numerical or approximate [12, 25, 31–33, 99, 103, 110, 129] results on the FPT problem already exist for specific classes of one dimensional stochastic processes. However, the case of bivariate processes has not been widely studied yet. Indeed, results are available only for specific problems such as the first exit time of a two-dimensional stochastic process from a specific surface [50, 66]. However there is a set of instances where the random variable of interest is the FPT of one of the components of a bivariate process through a constant or

a time dependent boundary. Beside the two-compartment neural model, described in Chapter 3, examples of this type of problems arise for the FPT of integrated processes such as the Integrated Brownian Motion (IBM) or the Integrated Ornstein Uhlenbeck Process (IOU). Indeed, these one dimensional stochastic processes should be studied as bivariate processes when the Markov property has to be preserved. Recent examples of applications of the IBM and the IOU processes appear in the metrological literature [89] where these processes are alternatively used to model the error of atomic clocks. In that case the crossing problem corresponds to the first attainment of an assigned value by the atomic clock error.

Motivated by these applications and by our interest on the ISIs generated by the two-compartment neural model, we consider the FPT of one component of a bivariate stochastic process through an assigned constant boundary. In particular we consider the class of bivariate Gauss-Markov processes, as their transition probability density function is well know.

In Section 4.1 we present a new integral equation for the FPT distribution and a condition for the existence and uniqueness of its solution. In Section 4.2 we introduce a numerical algorithm for its solution and we show its convergence properties. In Section 4.3 we illustrate the proposed numerical method through a set of examples. Finally in Section 4.4 we compare the computational effort and reliability of the proposed numerical method with a simulation algorithm.
We postpone to Chapter 5 the detailed application of the proposed numerical algorithm to the two-compartment neural model of Section 2.2.2.

4.1 An Integral Equation for the FPT distribution

Let us consider a diffusion process $\{\mathbf{X}(t), t \geq 0\}$ originated in $\mathbf{y} = \mathbf{0}$ at $t_0 = 0$.
Then the following theorem holds.

Theorem 4.1.1. *If*

$$\mathbb{P}\left(X_2(t) \geq S \,|X_1(\vartheta) = z, X_2(\vartheta) = S\right) \ , \ z \in \mathbb{R} \ , \ \vartheta \in [0, t] \tag{4.1}$$

and its derivative with respect to t are continuous in ϑ, then the FPT probability density function is the solution of the following integral equation

$$\begin{aligned}&\mathbb{P}\left(X_2(t) \geq S \,|\mathbf{X}(0) = \mathbf{0}\right) \\ &\qquad = \int_0^t d\vartheta \; g\left(\vartheta \,|\, \mathbf{0}, 0\right) \mathbb{E}_{Z(\vartheta)}\left[\mathbb{P}\left(X_2(t) \geq S \,|X_1(\vartheta), X_2(\vartheta) = S\right)\right]\end{aligned} \tag{4.2}$$

where the distribution of the random variable $Z(\vartheta)$ is the conditional distribution of $X_1(T)$ given $T = \vartheta$

$$\mathbb{P}\left(X_1(T) < z \,|T = \vartheta; \mathbf{X}(t_0) = \mathbf{y}\right). \tag{4.3}$$

The solution of (4.2) exists and it is unique.

Proof. Let us consider (1.43) with $x_2 = S$ and $x_1 = -\infty$. We get

$$\begin{aligned}&\mathbb{P}\left(X_2(t) > S \,|\, \mathbf{X}(0) = \mathbf{0}\right) \\ &\qquad = \int_0^t d\vartheta \int_{-\infty}^{+\infty} g_c\left((z, S), \vartheta \,|\, \mathbf{0}, 0\right) \mathbb{P}\left(X_2(t) > S | X_1(\vartheta) = z, X_2(\vartheta) = S\right) dz.\end{aligned} \tag{4.4}$$

Considering that

$$\begin{aligned}&\mathbb{P}\left(X_1(T) < z, T < t \;|\; \mathbf{X}(0) = \mathbf{0}\right) \\ &\qquad = \int_0^t d\tau \; \mathbb{P}\left(X_1(\tau) < z \,|T = \tau, \mathbf{X}(0) = \mathbf{0}\right) g\left(\tau \,|\mathbf{0}, 0\right),\end{aligned} \tag{4.5}$$

and taking the derivatives with respect to z and t, by using (1.42), we get

$$g_c\left((z, S), t \,|\mathbf{0}, 0\right) = \frac{\partial}{\partial z} \mathbb{P}\left(X_1(T) < z \,|T = t; \mathbf{X}(0) = \mathbf{0}\right) g\left(t \,|\mathbf{0}, 0\right). \tag{4.6}$$

Substituting (4.6) in (4.4) we get the integral equation (4.2). It is a first kind Volterra equation with regular kernel

$$k(t, \vartheta) = \mathbb{E}_{Z(\vartheta)}\left[\mathbb{P}\left(X_2(t) \geq S \,|X_1(\vartheta), X_2(\vartheta) = S\right)\right],$$

as $k(t, \vartheta)$ is bounded. In particular $k(t, t)$ does not vanish for any $t \geq 0$.
Due to hypothesis (4.1), the kernel of the Volterra equation (4.2) and its derivative with respect to t are continuous for $0 \leq \vartheta \leq t$.
Similarly, the left hand side of equation (4.2) and its derivative with respect to t are

continuous for $t \geq 0$. Furthermore $\mathbb{P}\left(X_2(0) \geq S \,|\mathbf{0}, 0\right) = 0$.
Thus, applying Theorem 5.1 of [76], we get the existence and uniqueness of the solution. □

Corollary 4.1.2. *The FPT probability density function of the second component of a bivariate Gauss-Markov process (1.25) satisfies the following equation*

$$1 - Erf\left(\frac{S - m^{(2)}(t)}{\sqrt{2Q^{(22)}(t)}}\right) = \tag{4.7}$$
$$= \int_0^t d\vartheta g\left(\vartheta \,|\mathbf{0}, 0\right) \mathbb{E}_{Z(\vartheta)}\left[1 - Erf\left(\frac{S - m^{(2)}(t \mid (X_1(\vartheta), S), \vartheta)}{\sqrt{2Q^{(22)}(t \mid (X_1(\vartheta), S), \vartheta)}}\right)\right],$$

where $m^{(2)}(t) = m^{(2)}(t \mid \mathbf{0}, 0)$ denotes the second component of the vector (1.28), $Q^{(22)}(t) = Q_t^{(22)}(t \mid \mathbf{0}, 0)$ denotes the element on the lower right corner of the matrix (1.29) and Erf(x) denotes the error function [1].
The Volterra equation (4.7) admits a unique solution if

$$\frac{\partial}{\partial t}\left(\frac{S - m^{(2)}(t \mid (z, S), \vartheta)}{\sqrt{2Q^{(22)}(t \mid (z, S), \vartheta)}}\right) \tag{4.8}$$

is a continuous function of $t \geq \vartheta \geq 0$.

Proof. Due to the Gaussianity of the process, we have

$$\begin{aligned}\mathbb{P}\left(X_2(t) \geq S \,|\mathbf{X}(t_0) = \mathbf{y}\right) &= \int_{-\infty}^{+\infty} dx_1 \int_S^{+\infty} dx_2 f\left(\mathbf{x}, t \mid \mathbf{y}, t_0\right) \\ &= \frac{1}{2}\left(1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t \mid \mathbf{y}, t_0)}{\sqrt{2Q^{(22)}(t \mid \mathbf{y}, t_0)}}\right)\right).\end{aligned}$$

Replacing this result into (4.2), we obtain (4.7).

Since (4.8) is continuous for hypothesis, then

$$\frac{\partial}{\partial t}\mathbb{P}\left(X_2(t) \geq S \mid \mathbf{X}(t_0) = \mathbf{y}\right) = \frac{\partial}{\partial t}\frac{1}{2}\left(1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t \mid \mathbf{y}, t_0)}{\sqrt{2Q^{(22)}(t \mid \mathbf{y}, t_0)}}\right)\right) \tag{4.9}$$
$$= -\frac{1}{\sqrt{\pi}} \exp\left\{-\left(\frac{S - m^{(2)}(t \mid \mathbf{y}, t_0)}{\sqrt{2Q^{(22)}(t \mid \mathbf{y}, t_0)}}\right)^2\right\} \frac{\partial}{\partial t}\left(\frac{S - m^{(2)}(t \mid \mathbf{y}, t_0)}{\sqrt{2Q^{(22)}(t \mid \mathbf{y}, t_0)}}\right)$$

is continuous. Thus applying Theorem 4.1.1 we get the existence and uniqueness of the solution of (4.7). □

Remark 4.1.3. The function

$$\mathbb{P}\left(X_2(t) \geq S \,|X_1(\vartheta) = z, X_2(\vartheta) = S\right) \tag{4.10}$$

represents the probability of being over the threshold S after a time interval $(t-\vartheta)$, starting from the threshold itself. For a Gauss-Markov process we get

$$\mathbb{P}\left(X_2(t) \geq S \,|X_1(\vartheta) = z, X_2(\vartheta) = S\right) = \frac{1}{2}\left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t \mid (z,S),\vartheta)}{\sqrt{2Q^{(22)}(t \mid (z,S),\vartheta)}}\right)\right]. \tag{4.11}$$

Thus under weak conditions of its well-definition, applying the l'Hopital's rule, the following limit

$$\lim_{\vartheta \to t}\left\{1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t \mid (z,S),\vartheta)}{\sqrt{2Q^{(22)}(t \mid \vartheta,(z,S))}}\right)\right\} \tag{4.12}$$

assumes a positive value $C \leq 2$. Then, by using the dominated convergence theorem, we can conclude that

$$\lim_{\vartheta \to t} \mathbb{E}_{Z(\vartheta)}\left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t \mid (X_1(\vartheta),S),\vartheta)}{\sqrt{2Q^{(22)}(t \mid (X_1(\vartheta),S),\vartheta)}}\right)\right] = C. \tag{4.13}$$

Remark 4.1.4. Note that the random variable $X_1(\vartheta)$ where $T = \vartheta$, in the expectation (4.13), has values on an interval $[a,b]$ that changes depending on the features of the process (1.25). Indeed it represents the position of the first component of the process when the second one has reached the boundary. For instance, in the IBM case, $X_1(\vartheta)$ has values in $[0,+\infty]$ as a negative value implies a negative increment in the second component, preventing a crossing of the boundary.

4.2 Gauss-Markov processes: a numerical algorithm

The complexity of equation (4.7) does not allow to get closed form solutions for g. Hence we pursue our study by introducing a numerical algorithm for its solution.

Let us consider the partition $t_0 = 0 < t_1 < \ldots < t_N = t$ of the time interval $[0, t]$ with step $h = t_k - t_{k-1}$ for $k = 1, \ldots, N$.
Discretizing the integral equation (4.7) via the Euler method, we have:

$$\begin{aligned} & 1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k)}{\sqrt{2Q^{(22)}(t_k)}}\right) \\ & = \sum_{j=1}^{k} \hat{g}\left(t_j \,|\mathbf{0}, 0\right) \mathbb{E}_{Z(t_j)} \left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k \mid (X_1(t_j), S), t_j)}{\sqrt{2Q^{(22)}(t_k \mid (X_1(t_j), S), t_j)}}\right)\right] h \end{aligned} \tag{4.14}$$

for $k = 1, \ldots, N$.

Equation (4.14) gives the following algorithm for the numerical approximation $\hat{g}\left(\tau \,|\mathbf{0}, 0\right)$ of $g\left(\tau \,|\mathbf{0}, 0\right)$, $\tau \in (0, t]$.

Step 1:

$$\hat{g}\left(t_1 \mid \mathbf{0}, 0\right) = \frac{1}{Ch}\left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_1)}{\sqrt{2Q^{(22)}(t_1)}}\right)\right], \tag{4.15}$$

where C is given by (4.13).

Step k, $k > 1$**:**

$$\begin{aligned} \hat{g}\left(t_k \mid \mathbf{0}, 0\right) &= \frac{1}{Ch}\left\{1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k)}{\sqrt{2Q^{(22)}(t_k)}}\right)\right\} \\ &- \frac{1}{C}\sum_{j=1}^{k-1} \hat{g}\left(t_j \mid \mathbf{0}, 0\right) \mathbb{E}_{Z(t_j)} \left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k \mid (X_1(t_j), S), t_j)}{\sqrt{2Q^{(22)}(t_k \mid (X_1(t_j), S), t_j)}}\right)\right]. \end{aligned} \tag{4.16}$$

Note that the first term on the r.h.s. is obtained for $j = k$.

To sum up, the FPT probability density function in the knots $t_0, t_1, \ldots, t_N$ is the solution of a linear system $\mathbf{L}\hat{\mathbf{g}} = \mathbf{b}$ where

$$\mathbf{b} = \begin{pmatrix} 1 - \mathrm{Erf}\left(\frac{S-m^{(2)}(t_1)}{\sqrt{2Q^{(22)}(t_1)}}\right) \\ \vdots \\ 1 - \mathrm{Erf}\left(\frac{S-m^{(2)}(t_N)}{\sqrt{2Q^{(22)}(t_N)}}\right) \end{pmatrix}, \quad \hat{\mathbf{g}} = \begin{pmatrix} \hat{g}\left(t_1 \mid \mathbf{0}, 0\right) \\ \vdots \\ \hat{g}\left(t_N \mid \mathbf{0}, 0\right) \end{pmatrix}$$

and

$$\mathbf{L} = \begin{pmatrix} Ch & & & & \\ \theta_{2,1}h & Ch & & & \\ \theta_{3,1}h & \theta_{3,2}h & Ch & & \\ \vdots & \vdots & & \ddots & \\ \theta_{N,1}h & \theta_{N,2}h & \cdots & \cdots & Ch \end{pmatrix},$$

with

$$\theta_{k,j} = \mathbb{E}_{Z(t_j)}\left[1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k \,|(X_1(t_j), S), t_j)}{\sqrt{2Q^{(22)}(t_k \,|(X_1(t_j), S), t_j)}}\right)\right] \tag{4.17}$$

for $k = 1, \ldots, N$ and $j = 1, \ldots, k$.

To evaluate the expected value (4.17) for $k = 1, \ldots, N$ and $j = 1, \ldots, k$, we make use of the following Monte Carlo method. We repeatedly simulate the bivariate process until the second component crosses the boundary and we collect the sequence $\{Z_i, i = 1, \ldots M\}$ of i.i.d random variables with probability distribution function (4.3). Then we compute the sample mean

$$\hat{\theta}_{k,j} = 1 - \frac{\sum_{i=1}^{M} \mathrm{Erf}\left(\frac{S-m^{(2)}(t_k|(Z_i,S),t_j)}{\sqrt{2Q^{(22)}(t_k|(Z_i,S),t_j)}}\right)}{M}. \tag{4.18}$$

Here M is the sample size.

The following lemma and theorem prove that this algorithm converges. In order to simplify the notations of the theorem, let us first define

$$\psi(z; t_k, t_j) := \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k \mid (z, S), t_j)}{\sqrt{2Q^{(22)}(t_k \mid (z, S), t_j)}}\right) \tag{4.19}$$

for $k = 1, \dots, N$, $j = 1, \dots, k$.

Lemma 4.2.1. *Let the parameters* $\mathbf{A}(t)$ *and* $\mathbf{M}(t)$ *of (1.25) be continuous on* $[0, t]$ *for each* $t_k, t_j \in [0, t]$ *and let the variance* $Q^{(22)}(t|(z, S), t_j)$ *be increasing with respect to* t*, then there exists a constant* γ*, such that for all* $h > 0$

$$\max_{1\le k\le N, 1\le j\le k-1} \mathbb{E}_{Z(t_j)} \left|\psi(Z(t_j), t_k, t_j) - \psi(Z(t_j), t_{k-1}, t_j)\right| \le \gamma h. \tag{4.20}$$

Proof. From the definition of $\psi(z; t_k, t_j)$, by using the increasing monotonicity of $Q^{(22)}(t|(z, S), t_j)$ we get

$$\begin{aligned}
|\psi(z, t_k, t_j) - \psi(z, t_{k-1}, t_j)| &= \frac{2}{\sqrt{\pi}} \left| \int_{\frac{S-m^{(2)}(t_{k-1}|(z,S),t_j)}{\sqrt{2Q^{(22)}(t_{k-1}|(z,S),t_j)}}}^{\frac{S-m^{(2)}(t_k|(z,S),t_j)}{\sqrt{2Q^{(22)}(t_k|(z,S),t_j)}}} e^{-y^2/2} dy \right| \\
&\le a_1 \left| \frac{S - m^{(2)}(t_k|(z, S), t_j)}{\sqrt{2Q^{(22)}(t_k|(z, S), t_j)}} - \frac{S - m^{(2)}(t_{k-1}|(z, S), t_j)}{\sqrt{2Q^{(22)}(t_{k-1}|(z, S), t_j)}} \right| \\
&\le \frac{a_1}{\sqrt{2Q^{(22)}(t_{k-1}|(z, S), t_j)}} \left| m^{(2)}(t_k|(z, S), t_j) - m^{(2)}(t_{k-1}|(z, S), t_j) \right|
\end{aligned} \tag{4.21}$$

where a_1 is a suitable constant.

From (1.28) we get

$$\begin{aligned}
&\left| m^{(2)}(t_k|\mathbf{y}, t_j) - m^{(2)}(t_{k-1}|\mathbf{y}, t_j) \right| \\
&\quad = |\phi(t_k, t_j) - \phi(t_{k-1}, t_j)|\, \mathbf{y} \\
&\qquad + \left| \phi(t_k, t_j) \int_{t_j}^{t_k} \phi(u, t_j)^{-1} \mathbf{M}(u) du - \phi(t_{k-1}, t_j) \int_{t_j}^{t_{k-1}} \phi(u, t_j)^{-1} \mathbf{M}(u) du \right| \\
&\quad = |\phi(t_k, t_j) - \phi(t_{k-1}, t_j)|\, \mathbf{y} \\
&\qquad + |\phi(t_k, t_j)| \left| \int_{t_j}^{t_k} \phi(u, t_j)^{-1} \mathbf{M}(u) du - \int_{t_j}^{t_{k-1}} \phi(u, t_j)^{-1} \mathbf{M}(u) du \right| \\
&\qquad + |\phi(t_k, t_j) - \phi(t_{k-1}, t_j)| \left| \int_{t_j}^{t_{k-1}} \phi(u, t_j)^{-1} \mathbf{M}(u) du \right|.
\end{aligned} \tag{4.22}$$

Since $\phi(t, t_j)$ is a continuous function of $t, t_j \in [0, t]$, it is globally bounded. Then it is Lipschitz with uniform Lipschitz constant with respect to t and t_j. Moreover, as $\phi(t, t_j)$ depends on $t_j \in [0, t]$ in a continuous way, it admits maximum on $[0, t]$. Therefore

$$|\mathbf{m}(t_k|\mathbf{y}, t_j) - \mathbf{m}(t_{k-1}|\mathbf{y}, t_j)| \leq K(t_j)h\mathbf{y} + O(h) \tag{4.23}$$

where $K(t_j)$ is a constant which depends on t_j.
By taking the expectation of (4.23) with $\mathbf{y} = (Z(t_j), S)$, we get

$$\mathbb{E}_{Z(t_j)} |\mathbf{m}(t_k|(Z(t_j), S), t_j) - \mathbf{m}(t_{k-1}|(Z(t_j), S), t_j)| = O(h). \tag{4.24}$$

Since $1/Q^{(22)}(t_{k-1}|(z, S), t_j)$ is bounded on $[0, t]$ for each z, the expectation of (4.21) gives

$$\mathbb{E}_{Z(t_j)} |\psi(Z(t_j), t_k, t_j) - \psi(Z(t_j), t_{k-1}, t_j)| = O(h) \tag{4.25}$$

and we get the thesis. □

Theorem 4.2.2. *If the sample size M for the Monte Carlo method is such that the error $|\lambda| = h^2$ at a confidence level α, then the error $\epsilon_k = \hat{g}\,(t_k \mid \mathbf{0}, 0) - g\,(t_k \mid \mathbf{0}, 0)$ of the proposed algorithm at the discretization knots t_k, $k = 1, 2, \ldots$, is $|\epsilon_k| = O(h)$ at the same confidence level α.*

Proof. Euler and Monte Carlo methods, applied to (4.7), give

$$1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k)}{\sqrt{2Q^{(22)}(t_k)}}\right) = \sum_{j=1}^{k} h\hat{g}\,(t_j\,|\mathbf{0}, 0)\,\hat{\theta}_{k,j}, \tag{4.26}$$

while (4.7) can be rewritten as

$$1 - \mathrm{Erf}\left(\frac{S - m^{(2)}(t_k)}{\sqrt{2Q^{(22)}(t_k)}}\right) = \sum_{j=1}^{k} hg\,(t_j\,|\mathbf{0}, 0)\left(\hat{\theta}_{k,j} + \lambda\right) + \delta(h, t_k) \tag{4.27}$$

where $\delta(h, t_k)$ denotes the error of Euler method and λ indicates the error of the Monte Carlo method at confidence level α.
Subtracting (4.27) from (4.26) we obtain

$$\delta(h, t_k) = h \sum_{j=1}^{k} \left(\hat{\theta}_{k,j}\epsilon_j + \lambda g\,(t_j\,|\mathbf{0}, 0)\right). \tag{4.28}$$

Differencing (4.28) and using (4.13), we get

$$\delta(h,t_k)-\delta(h,t_{k-1})=h\sum_{j=1}^{k-1}\left(\hat{\theta}_{k,j}-\hat{\theta}_{k-1,j}\right)\epsilon_j+hC\epsilon_k+g\left(t_k\left|\mathbf{0},0\right.\right)\lambda \tag{4.29}$$

or equally

$$\epsilon_k=\frac{1}{Ch}\left[\delta(h,t_k)-\delta(h,t_{k-1})\right]-\frac{1}{C}\sum_{j=1}^{k-1}\left(\hat{\theta}_{k,j}-\hat{\theta}_{k-1,j}\right)\epsilon_j-\frac{g\left(t_k\left|\mathbf{0},0\right.\right)\lambda}{hC}. \tag{4.30}$$

Then, due to Lemma (4.2.1) and to the law of large number, choosing M large enough, we have

$$\left|\epsilon_k\right|\leq\frac{1}{Ch}\left|\delta(h,t_k)-\delta(h,t_{k-1})\right|+\frac{ah}{C}\sum_{j=1}^{k-1}\epsilon_j+\frac{g\left(t_k\left|\mathbf{0},0\right.\right)\left|\lambda\right|}{hC}. \tag{4.31}$$

Finally, observing that the error of Euler method is $|\delta(h,t)|=O(h^2)$, choosing M such that the error of the Monte Carlo method is $|\lambda|=h^2$ and applying Theorem 7.1 of [76], we get the thesis. □

4.3 Examples

In this section we show the application of the algorithm presented in Sections 4.2 to two examples: an IBM and an IOU process.

4.3.1 Integrated Brownian Motion

The Integrated Brownian Motion by itself is not a Gauss-Markov diffusion process because it is not a Markov process. However we can study this one dimensional process as a bivariate process together with a standard Brownian motion, as follows

$$\begin{cases} dX_1(t) &= dB_t \\ dX_2(t) &= X_1(t)dt, \end{cases} \tag{4.32}$$

with $\mathbf{X}(0)=\mathbf{0}$.
The process (4.32) is a particular case of the Gauss-Markov diffusion process (1.25),

where

$$\mathbf{A}(t) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{M}(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{G}(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

There exist analytical solutions of the FPT problem of the integrated component of the process (4.32), but they are not efficient as they involve multiple integrals [55] or suppose particular symmetry properties [50].
Hence, we numerically solve the FPT problem for an IBM using the algorithm proposed in Section 4.2. A first attempt in this direction was discussed in [108]. In this instance the approximated FPT probability density function through a boundary S in the knots $t_0, t_1, \dots t_N$ is solution of a linear system $\mathbf{L}\hat{\mathbf{g}} = \mathbf{b}$ where

$$\mathbf{b} = \begin{pmatrix} 1 - \text{Erf}\left(\frac{\sqrt{6}S}{2t_1^{3/2}}\right) \\ \vdots \\ 1 - \text{Erf}\left(\frac{\sqrt{6}S}{2t_N^{3/2}}\right) \end{pmatrix}, \quad \hat{\mathbf{g}} = \begin{pmatrix} g\left(t_1 \mid \mathbf{0}, 0\right) \\ \vdots \\ g\left(t_N \mid \mathbf{0}, 0\right) \end{pmatrix}$$

and

$$\mathbf{L} = \begin{pmatrix} 2h & & & & \\ \theta_{2,1}h & 2h & & & \\ \theta_{3,1}h & \theta_{3,2}h & 2h & & \\ \vdots & \vdots & & \ddots & \\ \theta_{N,1}h & \theta_{N,2}h & \cdots & \cdots & 2h \end{pmatrix}.$$

Here $C = 2$, $Q^{(22)}(t_j) = t_j^3/6$ and for $k, j = 1, \dots, N$

$$\theta_{k,j} = \mathbb{E}_{Z(t_j)}\left[1 + \text{Erf}\left(\frac{\sqrt{6}X_1(t_j)}{2\sqrt{(t_k - t_j)h}}\right)\right].$$

Note that in this case the constant C defined in Remark 4.1.3 is equal to 2 and the range of the random variable $X_1(T)$ is $[0, \infty]$.
In Figure 4.1 we show the FPT probability density function of an IBM through a boundary S for three different values of the boundary.

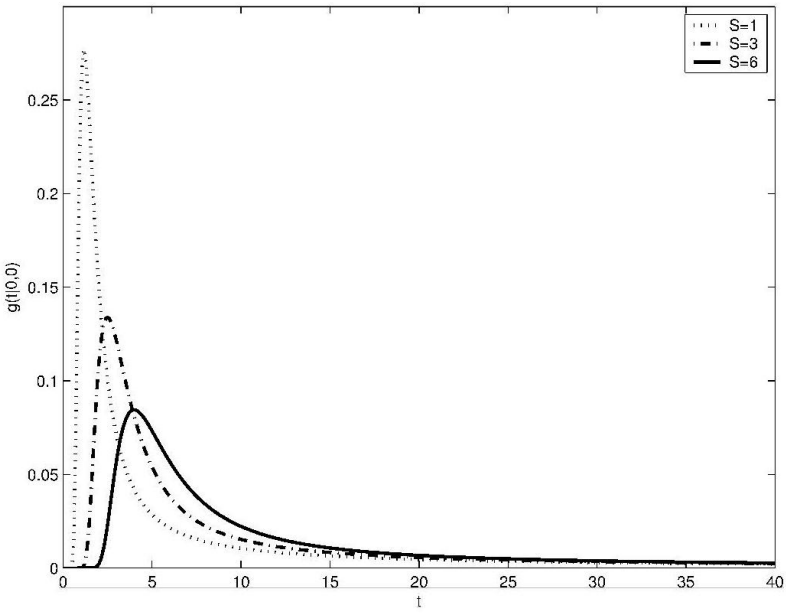


Figure 4.1: Evaluation of the FPT probability density function for an IBM through three different boundaries: $S = 1$ (dotted), $S = 3$ (dashdot), $S = 6$ (solid).

4.3.2 Integrated Ornstein Uhlenbeck Process

As the IBM, the IOU process is not a Markov process and it should be studied as a bivariate process together with an Ornstein-Uhlenbeck process, as follows

$$\begin{cases} dX_1(t) &= (-\alpha X_1(t) + \mu)\, dt + \sigma dB_t \\ \\ dX_2(t) &= X_1(t)dt, \end{cases} \tag{4.33}$$

with $\mathbf{X}(0) = \mathbf{0}$.

The process (4.33) is a Gauss-Markov diffusion process (1.25), where

$$\mathbf{A}(t) = \begin{pmatrix} -\alpha & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{M}(t) = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \text{ and } \mathbf{G}(t) = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that in this case the constant C defined in Remark 4.1.3 is equal to 1 and the range of the random variable $X_1(T)$ is $[-\infty, \infty]$.

In Figure 4.2 we show the FPT probability density function of an IOU through a boundary $S = 6$, for $\mu = 0.01$, $\sigma = 1$ and three different values of the parameter α.

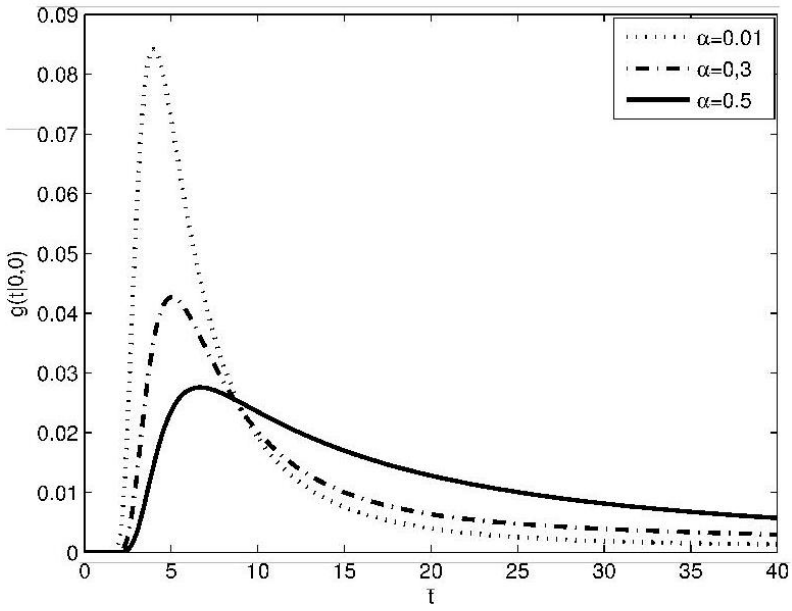


Figure 4.2: Evaluation of the FPT probability density function of an IOU through a boundary $S = 6$ for $\mu = 0.01$, $\sigma = 1$ and three different values of the parameter α: $\alpha = 0.01$ (dotted), $\alpha = 0.3$ (dashdot), and $\alpha = 0.5$ (solid).

Note that the curve for $\alpha = 0.01$ in Figure 4.2 is very similar to the curve for $S = 6$ in Figure 4.1. Indeed if $\mu \to 0$ and $\alpha \to 0$ an IOU converges to a standard IBM.

4.4 Comparison between the proposed numerical algorithm and simulation algorithms

The introduced numerical method involves a Monte Carlo estimation to evaluate the expected value (4.17). The trajectories are simulated by using an iterative expression of the solution of the stochastic differential equations (2.7a) and (2.7b), defining the

model. According to (1.26), in the autonomous case, this solution is

$$\mathbf{X}_t = e^{At}\mathbf{x}_0 + \mathbf{M}\int_0^t e^{A(t-s)}ds + \mathbf{J}_t, \tag{4.34}$$

where

$$\mathbf{J}_t = \int_0^t Ge^{A(t-s)}\mathbf{dB}_s.$$

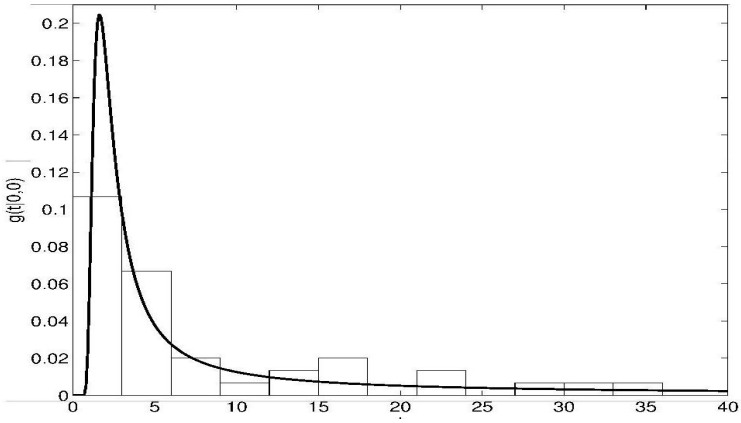
Here $\{\mathbf{J}_t, t \geq 0\}$, is a bivariate Gaussian process with zero mean and covariance matrix $\Sigma(t) = \mathbb{E}(\mathbf{J}_t \cdot \mathbf{J}_t')$, called innovation.
Taking a partition of the temporal interval $(0, t)$ with step $h = t_{n+1} - t_n$, we can obtain the following iterative form of equation (4.34)

$$\mathbf{X}_{t_{n+1}} = e^{Ah}\mathbf{X}_{t_n} + \mathbf{M}\int_{t_n}^{t_{n+1}} e^{A(t_{n+1}-s)}ds + \mathbf{J}_{t_n}. \tag{4.35}$$

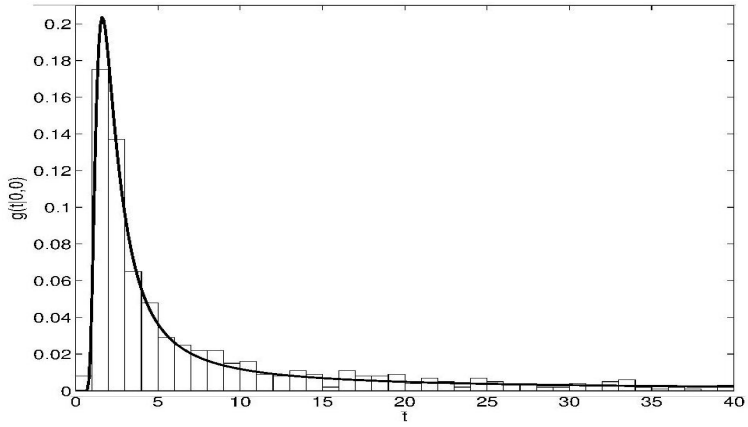
where $\mathbf{J}_{t_n} = \int_{t_n}^{t_{n+1}} Ge^{A(t_{n+1}-s)}\mathbf{dB}_s$.
The only term to compute now is the innovation $\mathbf{J}_{t_n}$. This is a simple simulation task, as $\mathbf{J}_{t_n}$ is a bivariate Gaussian process with zero mean and covariance matrix $\Sigma(t_n) = \mathbb{E}(\mathbf{J}_{t_n} \cdot \mathbf{J}_{t_n}')$, which depends only on the discretization step h.
Then, imposing on the second component a resetting mechanism after the generation of an action potential, we obtain our simulation algorithm. It provides trajectories which are not affected by errors. Furthermore we use a small discretization step in order to avoid hidden crossings.

One may wonder about the advantages of the proposed method compared to simulation algorithms. Indeed it is easy to simulate M sample paths of the considered bivariate process to get a sample of FPTs. It could be used to draw histograms or their continuous approximations. However this approach is computationally expensive. Indeed it requires large samples to give reliable results. Moreover the estimation of the distribution tails is scarcely reliable and time consuming.
On the contrary the proposed numerical method requests weak computational efforts, despite the presence of the Monte Carlo method. Indeed, a small number of trajectories guarantees reliable results. Moreover the efficiency of the numerical algorithm may be highlighted adding a stopping rule for the time window (or the number N

of knots) where the FPT probability density function is numerically evaluated. For instance, a common condition is to achieve a special value for the area under the FPT probability density function (e.g. 0.9), depending on the overall computational complexity.



(a) $M = 50$



(b) $M = 1000$

Figure 4.3: FPT probability density function for the IBM obtained via the proposed numerical method and the corresponding histogram. Samples of size M are used to build the histogram and to evaluate (4.18) in the numerical method: (a) $M = 50$ (b) $M = 1000$. Computational time for the numerical method: 2.0682 s (a) and 36.4803 s (b). Computational time for the histogram: 2.0798 s (a) and 36.5778 s (b).

In Figure 4.3 we compare the results obtained by the two methods. We simulate M sample paths of the IBM in order to determine a sample of M FPTs. We use it to draw the corresponding histogram. The same sample is used to compute the sample mean (4.18) to get the FPT probability density function via the numerical method. The choice $M = 1000$ gives reliable results in both cases. However, when $M = 50$ the histogram is crude while the numerical method does not lose its reliability. The computational times to build the histogram or to draw the FPT probability density function with the proposed numerical method are comparable, for the same value of M. The two methods run on Intel® Core™ i3-370M processor.

In Figure 4.4 we show the shapes of the FPT probability density function obtained via the proposed numerical algorithm. We use a sample of size $M = 50$ (solid line) and $M = 1000$ (dash line) to compute (4.18). Their differences are negligible.

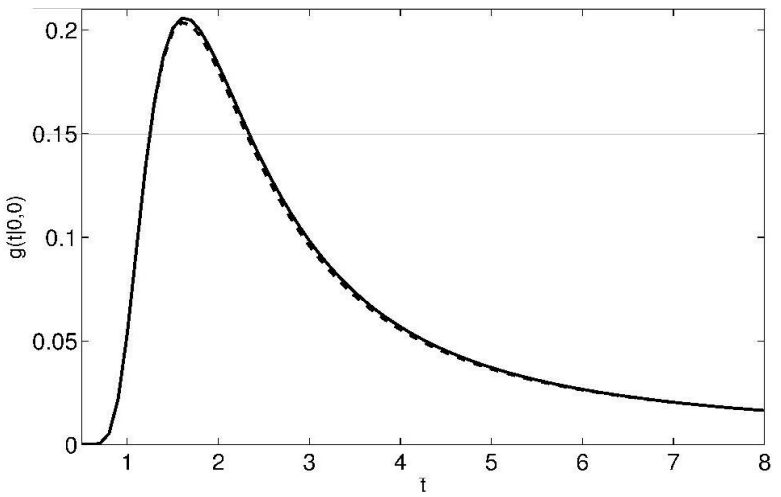


Figure 4.4: FPT probability density function for the IBM obtained via the proposed numerical method computed using a sample of size $M = 50$ (solid line) and $M = 1000$ (dash line) for the sample mean (4.18).

CHAPTER 5

The ISI distribution problem for a two-compartment neural model

Neural spike trains, described in Chapter 2, are typically stochastic in nature, due to the variability in the input they receive. As a consequence, how to read out accurately and efficiently the input information from spike trains is not a simple task. Indeed it is one of the central questions in theoretical neuroscience [36, 47, 101].
One of the prerequisites to acquire input information from a spike train is knowing the exact expression of ISI distribution.
Under the renewal hypothesis of LIF models, the ISI are i.i.d. random variables. Their distribution coincides with the distribution of the FPT of the one-dimensional stochastic process, modelling the neuron membrane potential, through an assigned

threshold. In this case there already exist some analytical, numerical and approximate results as mentioned in Chapter 4.
However the i.i.d. hypothesis is too strong to model neural data, as many experiments show the presence of an ISI dependency structure.

In Chapter 3 we show that the two-compartment neural model is able to reproduce dependent but identically distributed ISIs. It happens for particular choices of the parameters, when the dendritic component is stationary. In these instances, the ISI distribution coincides with the distribution of the FPT of the somatic component through a threshold S, when the model underlying process starts at $t_0 = t_{i^*}$. Here i^* denotes the index of the spiking epochs at which the dendritic component becomes stationary, as in Chapter 3.

Note that the two-compartment model underlying process, described by equations (2.7a) and (2.7b), is a Gauss-Markov diffusion process with

$$\mathbf{A}(t) = \begin{pmatrix} -\alpha - \alpha_r & \alpha_r \\ \alpha_r & -\alpha - \alpha_r \end{pmatrix}, \mathbf{M}(t) = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \text{ and } \mathbf{G}(t) = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}.$$

Then, we are able to approximate the probability density function of any ISI T_i, $i > i^*$, by applying the numerical algorithm proposed in Chapter 4.
Here the constant C defined in Remark 4.1.3 is equal to 2 and the range of the random variable $X_1(T)$ is $[kS, \infty]$, where

$$k = \frac{\alpha + \alpha_r}{\alpha_r}.$$

Indeed $X_1(T) < kS$ implies a negative increment in the second component, preventing a crossing of the boundary, as $dX_2(T) = \{-\alpha S + \alpha_r [X_1(T) - S]\}dt < 0$.

In Figure 5.1 we illustrate the ISI probability density function for different values of the model parameters α_r (panel a), σ (panel b) and μ (panel c). In each panel the other parameters are chosen to have identically distributed ISIs.
Note that these densities are not normal. Indeed they show slight asymmetries and

the normal assumption cannot be accepted (the p-values of normal goodness-of-fit tests are lower than 10^{-6}). This fact is evident in Figure 5.1(b), where the probability density functions with $\sigma = 5$ and $\sigma = 10$ are strongly asymmetric.

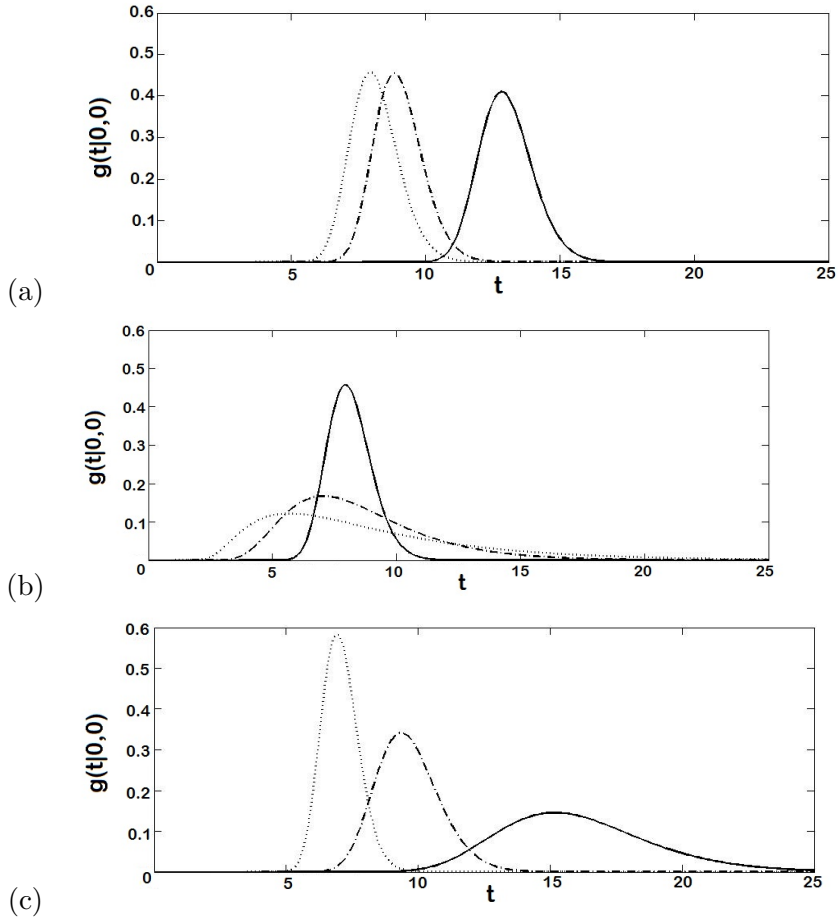


Figure 5.1: Probability density functions of identically distributed ISIs through a boundary $S = 10\ mV$, numerically computed by solving equation (1.44). In panel (a) $\alpha_r = 0.1\ ms^{-1}$ (solid), $\alpha_r = 0.3\ ms^{-1}$ (dashdot) and $\alpha_r = 0.5\ ms^{-1}$ (dotted). In panel (b) $\sigma = 1\ mV/ms^{1/2}$ (solid), $\sigma = 5\ mV/ms^{1/2}$ (dashdot) and $\sigma = 10\ mV/ms^{1/2}$ (dotted). In panel (c) $\mu = 2\ mV$ (solid), $\mu = 3\ mV$ (dashdot) and $\mu = 4\ mV$ (dotted). Furthermore $\alpha = 0.05\ ms^{-1}$, while $\alpha_r = 0.5\ ms^{-1}$ in (b) and (c), $\mu = 3.5\ mV$ in (a) and (b), $\sigma = 1\ mV/ms^{1/2}$ in (a) and (c).

When ISIs are dependent and identically distributed, they share the same distribution, approximated as shown before, and a particular dependency structure, described by a suitable copula. In these cases we can study the joint distribution of two successive ISIs, by applying Theorem 1.1.4.

5.1 Joint distribution of successive ISIs

In this Section we study the joint distribution of two successive dependent and identically distributed ISIs, by using some scatterplots of the associated copula for different values of μ (Figure 5.2).

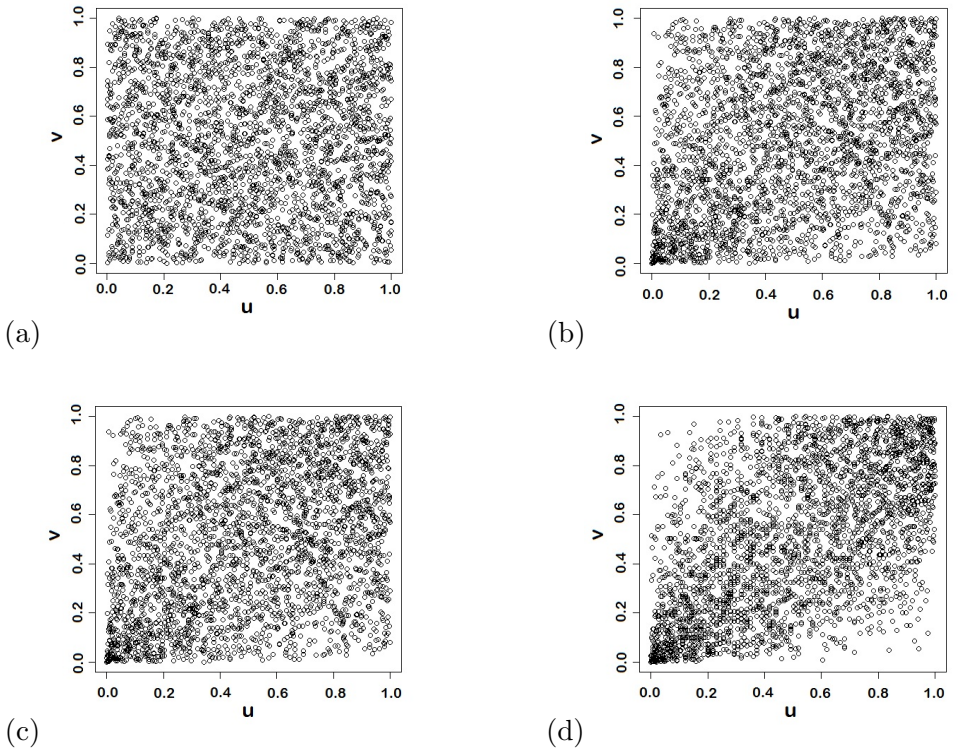


Figure 5.2: Scatterplot of the copula $C(u, v)$ between T_6 and T_5 for $\mu = 2\ mV$ (a), $\mu = 3.5\ mV$ (b), $\mu = 4\ mV$ (c), $\mu = 5\ mV$ (d).

The shape of these scatterplots allows us to hypothesize the presence of a normal copula. As μ increases, the scatterplots show a stronger dependence between subsequent ISIs, confirming the results of Table 3.4. The copula goodness of fit test, described in Section 1.1.2, confirms this conjecture, with p-values greater than 0.2. Hence the joint distribution of two subsequent ISIs can be obtained using the Gaussian copula, with correlation coefficient estimated from the data. The marginal distributions are obtained numerically with the techniques described in Chapter 4.
In Figure 5.3 we show an example of ISI joint probability density function.

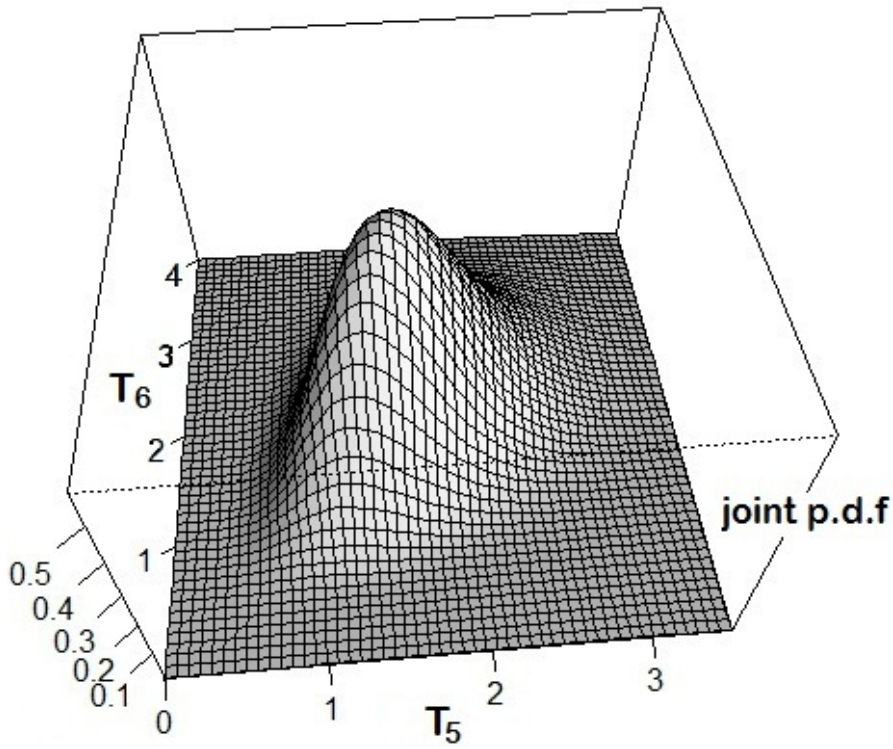


Figure 5.3: Numerical evaluation of the joint probability density function of T_5 and T_6 using a normal copula with correlation coefficient $\rho = 0.4$, estimated from data, and marginal distributions computed according to Chapter 5. The set of the parameters is: $\alpha = 0.05\ ms^{-1}$, $\alpha_r = 0.5\ ms^{-1}$, $\mu = 4\ mV$, $\sigma = 3\ mV/ms^{1/2}$, $S = 10\ mV$.

Similar results are obtained by varying other parameters when the successive ISIs are dependent but identically distributed.
It is useful to observe that the presence of a Gaussian copula between subsequent ISIs does not imply that the ISI marginals are normally distributed, as mentioned before. Indeed normal goodness of fit tests on ISI marginal distributions reject the Gaussian hypothesis with a p-value lower than 10^{-6} for any considered set of parameters.

CHAPTER 6

Firing rate estimators for a single spike train with dependent ISIs.

As already mentioned, one of the challenging problems of computational neuroscience concerns the features of the coding mechanisms used by neurons to encode external stimuli in sequences of action potentials [36, 47, 101].
In this chapter we focus our attention on a traditional coding scheme, which assumes that most of the information about the stimulus is contained in the proportion of spikes per time unit, the so-called firing rate.
Since the sequence of action potentials generated by a given stimulus to a neuron varies from trial to trial, neural firing rates are typically treated stochastically. The proportion of random events of the same kind per time unit is stochastically mod-

elled by the conditional intensity function (1.37) of a simple point process. Since the definition of conditional intensity function depends on the hazard rate functions (1.35a) and (1.35b), the firing rate estimation problem is strictly connected to the hazard rate estimation problem.

Estimators for point process hazard rate functions generally assume that the underlying point process is Poisson or a general renewal process, i.e. the inter-event intervals are i.i.d. random variables. These estimators have good statistical properties and converge quickly to the correct hazard rate functions as the sample size grows to infinity.
However the independence hypothesis is too strong for many applications, like in neuroscience. In the literature, the problem of dependence between inter-event intervals is addressed and solved only when their joint distribution is known [11, 44]. Indeed, there exists a number of examples of maximum likelihood estimators for a point process conditional and unconditional hazard rate functions, in presence of dependent inter-event intervals. Moreover it has been proven that the likelihood function of any sequence of inter-event intervals can be expressed in terms of the associated conditional and unconditional hazard rate functions ([21], Ch. 7). Nevertheless this type of approach is often not applicable to experimental instances in neuroscience, as the joint distribution of inter-event intervals is not always known in closed form. A modelling instance presenting this difficulty is determined by the ISIs generated by the two-compartment neural model, as discussed in Chapter 5.

In this chapter, we provide non-parametric estimators for the conditional and unconditional hazard rate functions of a simple point process with mildly dependent inter-event intervals. We assume in fact that the inter-event interval process is Markov, ergodic and stationary.

In Section 6.1 we describe the proposed non-parametric hazard rate estimators. Then we prove their uniform almost sure convergence to the unknown hazard rate function on a compact subset of $\mathbb{R}$. These hazard rate estimators are based on some hypotheses, which cannot be verified directly in many instances. Hence in Section 6.2 we

develop a statistical algorithm to validate our estimators on sample data. This "a posteriori" algorithm controls if the sample satisfies the necessary assumptions of the proposed method.
Then in Section 6.3 we illustrate the applicability of the proposed non-parametric estimators on a simple illustrative example, while in Section 6.4, we show their application to the neural firing rate estimation problem. In particular we apply the proposed estimators to a spike train simulated by the two-compartment neural model.

6.1 Uniform strongly consistent non-parametric estimators of a simple point process hazard rate functions

Let us consider a point process N on the time interval $[0, L]$, $0 < L < \infty$, with event-instant sequence $0 < l_1 < l_2 < \ldots < l_{N(L)}$. Assume that the inter-event intervals $T_i = l_i - l_{i-1}$, $i \geq 1$ and $l_0 = 0$, belong to a Markov, ergodic and stationary process. Hence the inter-event intervals are identically distributed with shared unconditional probability density function $f(t)$. Furthermore the marginal conditional probability density functions given the process history $f_i(t|\mathcal{F}_{i-1})$, defined in (1.36b), are $f_i(t|T_{i-1} = t_{i-1}) = f(t|t_{i-1})$, $i \geq 2$. Here $f(t|\tau)$ is a two variable transition probability density function, shared by all the inter-event intervals due to the stationarity assumption.

Then the hazard rate functions (1.35a) and (1.35b) of a point process N with Markov, ergodic and stationary inter-event intervals are

$$h(t) = \frac{f(t)}{S(t)} \,, \tag{6.1a}$$

$$h(t|\tau) = \frac{f(t|\tau)}{S(t|\tau)}. \tag{6.1b}$$

where $S(t) = 1 - \int_0^t f(s)ds$ and $S(t|\tau) = 1 - \int_0^t f(t|\tau)ds$ are the associated survival functions.

Thanks to the ergodic and stationary hypotheses, the sequence of inter-event intervals $\{T_1, T_2, \ldots, T_n\}$ satisfies the hypotheses of Theorem 1.3.4.
According to this theorem notation, let us denote by $\hat{f}_n(t)$ the uniform strongly consistent estimator of $f(t)$ on $[0, M]$, $M > 0$. Here M denotes the maximum value attainable for an inter-event interval of the point process N, observed on the time interval $[0, L]$. Hence $0 < M < L < \infty$ and the survival functions $S(t)$ and $S(t|\tau)$ are strictly greater than zero for every $t \in [0, M]$.

To prove our first main result we need the following auxiliary lemma.

Lemma 6.1.1. *Under the hypotheses of Theorem 1.3.4,*

$$\hat{S}_n(t) = 1 - \int_0^t \hat{f}_n(s)ds \ , \tag{6.2}$$

is a uniform strongly consistent estimator on $[0, M]$ *of the survival function* $S(t) = 1 - \int_0^t f(s)ds$*, that is*

$$\lim_{n\to+\infty} \sup_{t\in[0,M]} \left| S(t) - \hat{S}_n(t) \right| = 0 \ \textit{a.s.} \tag{6.3}$$

Proof.

$$\begin{aligned}
\sup_{t\in[0,M]} \left| S(t) - \hat{S}_n(t) \right| &= \sup_{t\in[0,M]} \left| \int_0^t \hat{f}_n(s)ds - \int_0^t f(s)ds \right| \\
&\leq \sup_{t\in[0,M]} \int_0^t \left| \hat{f}_n(s) - f(s) \right| ds \\
&\leq \sup_{t\in[0,M]} \int_0^t \sup_{s'\in[0,M]} \left| \hat{f}_n(s') - f(s') \right| ds \\
&= \sup_{s'\in[0,M]} \left| \hat{f}_n(s') - f(s') \right| \sup_{t\in[0,M]} \int_0^t ds \\
&= T \sup_{s'\in[0,M]} \left| \hat{f}_n(s') - f(s') \right| \ .
\end{aligned}$$

Finally, applying Theorem 1.3.4, we get the thesis. □

Remark 6.1.2. Observe that both $S(t)$ and $\hat{S}_n(t)$ are bounded and strictly positive functions on $[0, M]$, as they are, respectively, a survival function and a sum of survival functions associated to strictly positive density functions.

Now we have all the ingredients to define a uniform strongly consistent estimator of the unconditional hazard rate function (6.1a).

Proposition 6.1.3. *Under the hypotheses of Lemma 6.1.1,*

$$\hat{h}_n(t) = \frac{\hat{f}_n(t)}{\hat{S}_n(t)} \tag{6.4}$$

is a uniform strongly consistent estimator of the unconditional hazard rate function (6.1a) on $[0, M]$*, that is*

$$\lim_{n\to+\infty} \sup_{t\in[0,M]} \left|\hat{h}_n(t) - h(t)\right| = 0 \ \textit{a.s.} \tag{6.5}$$

Proof.

$$\begin{aligned}
\sup_{t\in[0,M]} \left|\hat{h}_n(t) - h(t)\right| &= \sup_{t\in[0,M]} \left|\frac{\hat{f}_n(t)S(t) - f(t)\hat{S}_n(t)}{\hat{S}_n(t)S(t)}\right| \\
&= \sup_{t\in[0,M]} \left|\frac{\hat{f}_n(t)S(t) - f(t)S(t) + f(t)S(t) - f(t)\hat{S}_n(t)}{\hat{S}_n(t)S(t)}\right| \\
&\leq \sup_{t\in[0,M]} \frac{\left|\hat{f}_n(t) - f(t)\right|}{\left|\hat{S}_n(t)S(t)\right|} |S(t)| + \sup_{t\in[0,M]} \frac{\left|S(t) - \hat{S}_n(t)\right|}{\left|\hat{S}_n(t)S(t)\right|} |f(t)| .
\end{aligned}$$

Applying Theorem 1.3.4 and Lemma 6.1.1 we get the thesis, as $S(t)$ and $\hat{S}_n(t)$ are bounded and strictly positive functions and $f(x) \in C_0(\mathbb{R})$. □

Let us now consider the estimation problem of the conditional hazard rate (6.1b). Due to the ergodic and stationary hypotheses, the sequence of couples of subsequent inter-event intervals $\{(T_i, T_{i+1}), i \geq 1\}$ satisfies the hypotheses of Theorem 1.3.6. Hence it holds:

Proposition 6.1.4. *Consider a uniform strongly consistent kernel estimator $\hat{f}_n(t|\tau)$ on $[0, M]^2$ of the inter-event interval conditional density function $f(t|\tau)$, according to Theorem 1.3.6. Furthermore let the involved kernels be both strictly positive on $\mathbb{R}$. Then*

$$\hat{h}_n(t|\tau) = \frac{\hat{f}_n(t|\tau)}{\hat{S}_n(t|\tau)} \ , \tag{6.6}$$

is a uniform strongly consistent estimator of the conditional hazard rate function (6.1b) on $[0, M]^2$, that is

$$\lim_{n\to+\infty} \sup_{(\tau,t)\in[0,M]^2} \left|\hat{h}_n(t|\tau) - h(t|\tau)\right| = 0 \ a.s. \tag{6.7}$$

Proof. The proof follows the same line of Proposition 6.1.3. □

Remark 6.1.5. Using Proposition 6.1.3 and Proposition 6.1.4, we can also provide a uniform strongly consistent estimator for the conditional intensity function (1.37) of a Markov, ergodic and stationary inter-event interval process, that is

$$\hat{\lambda}^*_n(t) = \begin{cases} \hat{h}_n(t) & (0 < t \leq l_1) \\ \hat{h}_n(t - l_{j-1}|T_{j-1} = t_{j-1}) & (l_{j-1} < t \leq l_j, j = 2, \dots n) \end{cases} \tag{6.8}$$

where $l_1, l_2, \dots, l_n$ are the event instants.

Remark 6.1.6. Modelling spike trains by means of simple point processes, the conditional intensity function estimator (6.8) can be applied as firing rate estimator for neurons generating statistically dependent but identically distributed ISIs.

6.2 A statistical algorithm to validate the proposed hazard rate estimators on sample data

The proposed hazard rate estimators are based on Markov, ergodic and stationary hypotheses. Therefore when these hypotheses cannot be verified directly, we need to validate the proposed estimators on sample data. In this section we present an "a posteriori" testing procedure, based on Theorem 1.2.9, to verify whether the estimators are reliable as well as consistent.

We first compute the conditional intensity function estimator $\hat{\lambda}_n^*$, defined by equation (6.8). Then we perform the following time transformation

$$t \mapsto \hat{\Lambda}_n^*(t) = \int_0^t \hat{\lambda}_n^*(u)du. \tag{6.9}$$

Under this time rescaling, the inter-event intervals $T_i = l_i - l_{i-1}$, $i \geq 1$ and $l_0 = 0$, become

$$\tilde{T}_i = \hat{\Lambda}_n^*(l_i) - \hat{\Lambda}_n^*(l_{i-1}) = \int_{l_{i-1}}^{l_i} \hat{\lambda}_n^*(u)du, \quad i = 1, \ldots, n. \tag{6.10}$$

If the hazard rate estimators (6.4) and (6.6) are reliable, i.e. if the hypotheses supporting their computation are verified, the transformed inter-event intervals $\tilde{T}_i$, $i = 1, \ldots, n$, should be i.i.d. exponential random variables with mean 1, according to Theorem 1.2.9. Hence a way to validate our estimators on sample data is to check the independence and the exponential distribution of these transformed inter-event intervals.
For this purpose, we can simply perform a goodness-of-fit test to verify if the random sequence $\{\tilde{T}_i, i = 1, \ldots, n\}$ follows the exponential distribution with mean 1. Then we can compute the Kendall's tau of the couples $(\tilde{T}_i, \tilde{T}_{i+1})$, $i = 1, \ldots, n-1$, to check if the transformed inter-event intervals are independent.

Here we propose an alternative test of the hypotheses on the transformed inter-event intervals, based on the concept of independent copula (1.9). Indeed, under the null hypothesis, the bivariate copula between two subsequent inter-event intervals should be the independent copula, with exponential marginal distributions of mean 1.
Actually we consider a further transformation of the inter-event intervals:

$$Z_i = 1 - e^{-\tilde{T}_i} \ . \tag{6.11}$$

Under the null hypothesis, $\{Z_i, i \geq 1\}$ is a collection of i.i.d. uniform random variables on the interval $[0, 1]$. Therefore the copula of the couples (Z_i, Z_{i+1}), $i = 1, 2, \ldots, n$, should be the independent copula with uniform marginal distributions on $[0, 1]$. Hence we can test the reliability of the proposed estimators, performing a uniformity test and the goodness-of-fit test for the independent copula proposed in Section

1.1.2. Here the approximate p-values for the copula goodness-of-fit test statistic are obtained using a parametric bootstrap, described in [42] and [43].

Therefore an "a posteriori" validation procedure for the proposed estimators of the unconditional and conditional hazard rate functions is based on the following statistical algorithm.

Algorithm 6.2.1. (*Validation algorithm*)

1. Construct the non-parametric estimator (6.8).
2. Compute the transformed inter-event intervals
$$\tilde{T}_i = \int_{l_{i-1}}^{l_i} \hat{\lambda}_n^*(u)du \ , \ i = 1, 2, \ldots, n.$$
3. Test the hypothesis that $\tilde{T}_i$ are i.i.d. exponential variables with mean 1:
   a. Make the transformation $Z_i = 1 - e^{-\tilde{T}_i}$, $i = 1, 2, \ldots, n$ and test whether these transformed random variables are uniform on $[0, 1]$.
   b. Perform a goodness-of-fit test for the independent copula on the couples (Z_i, Z_{i+1}), $i = 1, 2, \ldots, n-1$.

Remark 6.2.2. Testing that Z_i, $i = 1, 2, \ldots, n$, are i.i.d. uniform random variables on $[0, 1]$ is equivalent to test that $\tilde{T}_i$, $i = 1, 2, \ldots, n$, are i.i.d. exponential random variables with mean 1.

Remark 6.2.3. In Algorithm 6.2.1 we perform a goodness-of-fit test for copulas to verify the independence between Z_i and Z_{i+1}. However any other test of independence can be applied, like a classical chi-squared test.

When the inter-event intervals are not Markov, extensions of the proposed estimators to the case of stochastic processes with finite memory are needed. Indeed the original version of Theorem 1.3.6 in [86] and [4] concerns the estimation of the conditional probability density function of a univariate random variable Y given a generic

d-variate random variable X, $d \geq 1$. Therefore the extension of the proposed estimators to inter-event interval processes with finite memory is immediate, although this implies stronger computational efforts and requests larger samples.

On the other side, when the inter-event interval process is neither ergodic nor stationary, we need some strategy to ensure the identical distribution of inter-event intervals. For instance we can divide the estimation problem on shorter time windows on which the process is statistically stationary and ergodic. Otherwise if the process is periodic and we know its period, we can sample the inter-events intervals spaced by a period to ensure their identical distribution.

6.3 A simple illustrative example

In this section we use simulated data to show the features of the non-parametric hazard rate estimators presented in Section 6.1. We simulate a sample of inter-event intervals from a non-negative and ergodic autoregressive (AR) model of order 1, and we prove the efficacy of the validation procedure of Section 6.2.

Definition 6.3.1. A *non negative AR model of order 1* is defined as

$$X_k = \phi X_{k-1} + \xi_k, \quad k \in \mathbb{N}, \tag{6.12}$$

where $X_0 = 0$, ϕ is a non-negative parameter and ξ_k, $k \in \mathbb{N}$, are i.i.d. non-negative random variables.

Remark 6.3.2. As a classical Gaussian AR model of order 1 [114], a non-negative AR model of order 1 is Markov. Furthermore, if we choose $0 < \phi < 1$, it is also ergodic and stationary.

Remark 6.3.3. Since ϕ and ξ_k, are both non-negative, also X_k is non-negative for every $k > 0$. Hence a non-negative AR model of order 1 can simulate properly a sequence of inter-event intervals.

Here we simulate 1000 inter-event intervals from a non-negative AR model of order 1 with ξ_k, $k \in \mathbb{N}$, exponentially distributed with mean 1.
To compute the hazard rate estimators (6.4) and (6.6) we choose

1. Gaussian kernels K_1 and K_2 with mean zero and standard deviation $\sigma = 0.3$;
2. Kernel weights $b_n = n^{-\beta}$, where n is the sample size and $\beta = 0.2$.

Remark 6.3.4. Note that the kernels and their weights satisfy the hypotheses of Theorem 6.1.3 and Theorem 6.1.4.

In Table 6.1 we report the results of the validation Algorithm 6.2.1 proposed in Section 6.2 for different choices of the parameter ϕ. The table shows the p-values for the uniformity test and the copula goodness-of-fit test, performed in the second step of the algorithm.

ϕ	**Uniformity test p-value**	**Copula goodness-of-fit test p-value**
0.2	0.67	0.79
0.5	0.60	0.92
0.8	0.40	0.84
1	10^{-4}	10^{-16}
1.5	10^{-4}	10^{-16}

Table 6.1: Results of the validation Algorithm 6.2.1, applied on inter-event intervals simulated by a non-negative AR model of order 1.

When $0 < \phi < 1$ both tests correctly return high p-values, as the ergodic and stationary hypotheses are verified (first three lines of Table 6.1). On the other hand, when $\phi \geq 1$, the inter-event intervals are not stationary. As a result, the validation algorithm returns low p-values for both the uniformity test and the copula goodness-of-fit test (last two lines of Table 6.1). Hence, in these cases, our validation algorithm correctly alerts on the wrong use of the proposed estimators (6.4) and (6.6).

6.4 Application to the firing rate estimation problem.

Aim of the neural code is to transform the input stimulus into a neuronal response. A traditional coding scheme, called rate coding, assumes that most information about the stimulus is contained in the proportion of action potentials per time unit, the so-called firing rate. Existing estimators request the independence and the identically distribution of ISIs. However, as already outlined in previous chapters, the hypothesis of independence is too strong for neural data.

Typically spike trains are modelled by stochastic simple point processes. Indeed spike trains are stochastic sequences of events of the same kinds, which admit at most one single spike at any time instant. In particular, here we consider simple point processes with dependent inter-event intervals, whose conditional intensity function is the modelling counterpart of the firing rate of a neuron with dependent ISIs.

In this section we estimate the firing rate of a neuron modelled by the two-compartment model, described in Section 2.2.2. Indeed the ISIs generated by this model can be statistically stationary and Markov for particular choices of the parameter, as analysed in Chapter 3. Therefore they constitute an appropriate sample on which to apply the conditional intensity function estimator (6.8).

Here we simulate 1000 such stationary and Markov ISIs for suitable choices of the model parameters. Then we compute the conditional intensity function estimator (6.8) using

1. Gaussian kernels K_1 and K_2 with mean zero and standard deviation $\sigma = 0.2$;
2. Kernel weights $b_n = n^{-\beta}$ where n is the sample size and $\beta = 0.2$.

Remark 6.4.1. Note that the kernels and their weights satisfy the hypotheses of Theorem 6.1.3 and Theorem 6.1.4.

In Table 6.2 we report the results of the validation Algorithm 6.2.1 proposed in section 6.2 for different choices of the model parameters, such that the ISIs are stationary and Markov. It shows the p-values for the uniformity test and the copula

goodness-of-fit test, performed in the second step of the algorithm.

Parameters	**Uniformity test p-value**	**Copula goodness-of-fit test p-value**
$\alpha = 0.05$, $\alpha_r = 0.5$, $\mu = 4$, $\sigma = 1$, $S = 10$	0.88	0.97
$\alpha = 0.05$, $\alpha_r = 0.5$, $\mu = 3.5$, $\sigma = 5$, $S = 10$	0.65	0.49
$\alpha = 0.05$, $\alpha_r = 0.25$, $\mu = 4$, $\sigma = 1$, $S = 10$	0.84	0.62
$\alpha = 0.05$, $\alpha_r = 0.5$, $\mu = 8$, $\sigma = 1$, $S = 10$	0.21	0.01

Table 6.2: Results of the validation Algorithm 6.2.1, applied on ISIs simulated by the two-compartment neural model.

Remark 6.4.2. When the input intensity μ is close to the value of the firing threshold S, the ISIs become very short and the evolution of the two-compartment neural model is more dependent on its past history. For these instances the copula goodness-of-fit test correctly fails (last line of Table 6.2), as the ISI process is statistically a Markov process of order greater than 1 (see Table 3.5).

Conclusion

Motivated by our interest in neural modelling, in this thesis we develop and investigate with different techniques of stochastic and statistical analysis the dependency structures, which can arise in neural information processing.

From a stochastic modelling viewpoint, we consider a particular bivariate stochastic model, called two-compartment neural model, which is able to reproduce the dependency structures observed on experimental data. Unlike the classical and more popular one-dimensional LIF models, this model attempts to account for the geometry of the nerve cell. Hence it allows the investigation of the effects of the interaction between different parts of the neuron.
Of course we do not claim that the use of a bivariate stochastic process to model the neural dynamics makes this model more realistic. For instance, this model assumes that the compartments are infinitely close to one other and disregards the existence of other compartments. Further constraints should be introduced to make the model biologically more acceptable. We are interested in this model as it seems to be one of the simplest models allowing a serial dependence between successive ISIs.

Despite the increased complexity of the two-compartment neural model, we are able

to develop some suitable mathematical tools for its analysis. Furthermore, in this thesis we make use of some dependency measures, like copulas and the Kendall's tau, which are not common in neuroscience. In Chapters 3 and 5 we show their power and we suggest their regular use on recorded data.

Then, we develop further statistical methods for the analysis of two fundamental aspects of the neural information processing: the ISI distribution and the firing rate. In Chapters 4 and 5 we provide a numerical algorithm for the estimation of the ISI distribution as the FPT distribution of one component of the bivariate stochastic process, underlying the two-compartment model. We show that it is solution of a new integral equation and we prove its existence and uniqueness. We also discuss the advantages of this numerical method with respect to a totally simulated algorithm. Chapter 6 presents a contribution to the improvement of the class of neural firing rate estimators. Existing estimators request the independence of available data. Here we propose a non-parametric estimator in case of Markov, ergodic and stationary ISI processes. We also provide a statistical algorithm to validate it on sample data, when the hypotheses on the ISIs cannot be verified directly. Extensions to stochastic processes with finite memory are briefly discussed.

The results presented in this thesis are motivated by neuroscience problems. However their interest is wider and applications in different contexts can arise. For instance multi-compartment models can be applied to model pharmacokinetics, ecosystems and computer networks. Indeed the presented two-compartment model can be generalised to dimension d, $d > 2$, adding a finite number of further compartments. However this generalization requires more mathematical efforts to generalize the numerical method for the estimation of the FPT distribution.

The hazard rate estimators, proposed in Chapter 6, to estimate the neural firing rate can be applied in many fields like reliability theory, epidemiology and economics. Future extensions of the range of applicability of these estimators concern the development of statistical techniques devoted to isolate periodic changes or trends of the underlying stochastic processes, as proposed in [59].

Bibliography

[1] Abramowitz M. and Stegun I.A. (1964) *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, Dover, New York.

[2] Adrian E.D. and Zotterman Y. (1926) *The impulses produced by sensory nerve-endings*, The Journal of physiology, 61(2), 151-171.

[3] Adrian E.D. (1928) *The basis of sensation*, W.W. Norton, New York.

[4] Arfi M. (1998) *Nonparametric prediction from ergodic samples*, Journal of Nonparametric Statistics 9, 23-37.

[5] Arnold L. (1974) *Stochastic differential equations: theory and applications*, Krieger publishing company, Malabar, Florida.

[6] Benedetto E., Sacerdote L. and Zucca C. (2013) *A first passage problem for a bivariate diffusion process: Numerical solution with an application to neuroscience when the process is Gauss-Markov*, J. Computational Applied Mathematics 242, 41-52.

[7] Benedetto E. and Sacerdote L. (2013) *On dependency properties of the ISIs generated by a two-compartmental neuronal model*, Biological cybernetics, 107(1), 95-106.

[8] Berman M. (1983) *Comment on "Likelihood analysis of point processes and its applications to seismological data" by Ogata*, Bulletin Internat. Stat. Instit. 50, 412-418.

[9] Boente G. and Fraiman R. (1988) *Consistency of a non-parametric estimate of a density function for dependent variables*, J. Multivariate Anal. 25, 90-99.

[10] Bressloff P.C. (1995) *Dynamics of a compartmental integrate-and-fire neuron without dendritic potential reset*, Phys. D. 80, 399-412.

[11] Brown E.N., Barbieri R., Ventura V., Kass R.E., Frank L.M. (2001) *The time rescaling theorem and its application to neural spike train data analysis*, Neural Computation 14, 325-346.

[12] Buonocore A., Nobile A.G. and Ricciardi L.M. (1987) *A new integral equation for the evaluation of first-passage-time probability densities*, Adv. in Appl. Probab. 19 4, 784-800.

[13] Buonocore A., Caputo L., Pirozzi E. and Ricciardi L. M. (2010) *On a Stochastic Leaky Integrate-and-Fire NeuronalModel*, Neural Computation 22, 2558-2585.

[14] Burkitt A.N. (2006) *A review of the integrate and fire neuron model: I. Homogeneous synaptic input*, Biol. Cybern. 95, 1-19.

[15] Burkitt A.N. (2006) *A review of the integrate and fire neuron model: II. Inhomogeneous synaptic input and network properties*, Biol. Cybern. 95, 97-112.

[16] Capocelli R.M. and Ricciardi L.M. (1971) *Diffusion approximation and first passage time problem for a model neuron*, Kybernetik 8, 214223.

[17] Cocatre-Zilgien J.H. and Delcomyn F. (1992) *Identification of bursts in spike trains*, J. Neurosci. Meth. 41, 19-30.

[18] Cox D.D.R. and Isham V. (1980) *Point processes*, Monographs on Applied Probability and Statistics, Chapman & Hall, London-New York.

[19] Cramér H. (1928) *On the composition of elementary errors*, Skand. Aktuarietids, Vol. 11, 13-74 and 141-180.

[20] Cullheim S. and Kellerth J.O. (1978) *A morphological study of the axons and recurrent axon collaterals of cat sciatic alpha-motoneurons after intracellular staining with horseradish peroxidase*, J Comp Neurol. 178(3), 537-57.

[21] Daley D.J. and Vere-Jones D. (2008) *An Introduction to the Theory of Point Processes, Vol. 2: General Theory and Structure (2_nd edition)*, Springer, New York.

[22] De la Prida L.M., Stollenwerk N. and Sanchez-Andres J.V. (1997) *Bursting as a source for predictability in biological neural network activity*, Physica D 110, 323-331.

[23] Delacroix M., Nogueira M.E. and Martins A.C. (1992) *Sur l'estimation de la densit d'observations ergodiques*, Stat. et Anal. Donnes. 16, 25-38.

[24] Darling D.A. and Siegert A.J.F. (1953) *The first passage problem for a continuous Markov process*, Ann. Math. Stat. 24, 624-639.

[25] Di Nardo E., Nobile A., Pirozzi E. and Ricciardi L.M. (2001) *A computational approach to first-passage-time problems for Gauss-Markov processes*, Adv. Appl. Prob. 33, 453-482.

[26] Ditlevsen S. and Lansky P. (2007) *Parameters of stochastic diffusion processes estimated from observations of first hitting times: Application to the leaky integrate-and-fire neuronal model*, Physical Review E, 76:41096.

[27] Ditlevsen S. and Ditlevsen O. (2008) *Parameter estimation from observations of first passage times of the Ornstein-Uhlenbeck process and the Feller process*, Probabilistic Engineering Mechanics, 23, 170-179.

[28] Ditlevsen S. and Lansky P. (2008) *Comparison of statistical methods for estimation of the input parameters of the Ornstein-Uhlenbeck neuronal model from first passage time data*, In Collective Dynamics: Topics on Competition and Cooperation in the Bioscience, Volume CP1028 of American Institute of Physics Proceedings Series, L.M. Ricciardi, A. Buonocore, E. Pirozzi (Ed.), American Institute of Physics, Melville, NY, 171-185.

[29] Ditlevsen S. and Greenwood P. (2012) *The Morris-Lecar neuron model embeds a leaky integrate-and-fire model*, J math Biol, 1-21.

[30] Doob J.L. (1953) *Stochastic processes*, John Wiley and Sons, New York.

[31] Downes A.N. and Borovk K. (2008) *First Passage Densities and Boundary Crossing Probabilities for Diffusion Processes*, Methodology and Computing in Applied Probability, Volume 10, Number 4, 621-644.

[32] Durbin J. (1985) *The first passage density of a continuous Gaussian process to a general boundary*, J.Appl. Prob. 22, 99-122.

[33] Durbin J. (1992) *The first passage density of the Brownian motion process to a curved boundary*, J. Appl. Prob. 29, 291-304.

[34] Engel T. A., Schimansky-Geier L., Herz A. V., Schreiber S. and Erchova I. (2008) *Subthreshold membrane-potential resonances shape spike-train patterns in the entorhinal cortex*, Journal of neurophysiology, 100, 1576-1589.

[35] Farkhooi F., Strube-Bloss M.F. and Nawrot M.P. (2009) *Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability*, Physical Review E 79(2).

[36] Feng J. (2003) *Computational neuroscience: a comprehensive approach*, CRC press.

[37] Fermanian J.D. (2005) *Goodness-of-fit tests for copulas*, Journal of Multivariate Analysis 95, 119-152.

[38] Fredricks G.A. and Nelsen R.B. (2007) *On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables*, Journal of Statistical Planning and Inference.

[39] Frees E.W. and Valdez E.A. (1998) *Understanding relationships using copulas*, North American Actuarial Journal 2(1), 1-26.

[40] Folland G.B. (1999) *Real Analysis: modern techniques and their applications*, John Wiley and Sons, New York.

[41] Genest C. and Mackay J. (1986) *The joy of copulas: Bivariate distributions with uniform marginals*, The American Statistician 40, 280-283.

[42] Genest C. and Rémillard B. (2008) *Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models*, Annales de lInstitut Henri Poincaré: Probabilités et Statistiques, Vol. 44, No. 6, 1096-1127.

[43] Genest C., Remillard B. and Beaudoin D. (2009) *Goodness-of-fit tests for copulas: A review and a power study*, Insurance: Mathematics and Economics 44, 199-214.

[44] Gerhard F., Haslinger R., Pipa G. (2011) *Applying the multivariate time-rescaling theorem to neural population models*, Neural comput. 23, 1452-1483.

[45] Gerstein G.L. and Mandelbrot B. (1964) *Random walk models for the spike activity of a single neuron*, Biophys. J. 4, 41-68.

[46] Gerstner W., Kreiter A. K., Markram H. and Herz A. V. (1997) *Neural codes: firing rates and beyond*, Proceedings of the National Academy of Sciences, 94(24), 12740-12741.

[47] Gerstner W. and Kistler W.K. (2002) *Spiking Neuron models*, Cambridge University Press.

[48] Giorno V., Nobile A.G. and Ricciardi L.M. (1989) *A symmetry-based constructive approach to probability densities for one-dimensional diffusion processes*, J. Appl. Prob. 27, 707-721.

[49] Giorno V., Nobile A.G. and Ricciardi L.M. (1990) *On the asymptotic behavior of first-passage-time densities for one dimensional diffusion processes and varying boundary*, Adv. Appl. Prob. 22, 883-914.

[50] Giorno V., Nobile A.G., Ricciardi L.M. and Di Crescenzo A. (1995) *On a symmetry-based constructive approach to probability densities for two-dimensional diffusion processes*, J. Appl. Prob. 32, 316-336.

[51] Giraudo M.T., Greenwood P. and Sacerdote L. (2011) *How sample paths of leaky integrate-and-fir models are influenced by the presence of a firing threshold*, Neural Comput. 23, no. 7, 1743-1767.

[52] Giraudo M.T., Sacerdote L. and Sirovich R. (2013) *Nonparametric Estimation of Mutual Information through the Entropy of the Linkage*, Entropy (preprint).

[53] Glaser E.M. and Van der Loos H. (1965) *A semiautomatic computer microscope for the analysis of neuronal morphology*, IEEE Trans Biomed Eng. (12), 22-31.

[54] Godfrey K. (1983) *Compartmental models and their application*, Academic Press, New York.

[55] Goldman M. (1971) *On the First Passage of the Integrated Weiner Process*, The Annals of Mathematical Statistics, Vol. 42, No. 6, 2150–2155.

[56] Gyorfi L. (1981) *Strong consistent density estimate*, Journal of multivariate analysis 11, 81-84.

[57] Hodgkin A.L. and Huxley A.F. (1952) *A quantitative description of membrane current and its application to conduction and excitation in nerve*, J. Physiol. 117(4), 500-544.

[58] Hubel D.H. and Wiesel T.N. (1959) *Receptive fields of single neurons in the cat's striate cortex*, The Journal of physiology, 148(3), 574-591.

[59] Iolov A., Ditlevsen S. and Longtin A. (2013) *Fokker-Planck and Fortet equation-based parameter estimation for a leaky integrate-and-fire model with sinusoidal and stochastic forcing*, To appear in Journal of Mathematical Neuroscience.

[60] Izenman A.J. (1991) *Recent developments in non-parametric density estimation*, J. Amer. Statist. Assoc. 86, 205-224.

[61] Joe H. (1997) *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[62] Karlin S., Taylor H.M. (1975) *A first course in stochastic processes*, Academy press, New York.

[63] Karlin S., Taylor H.M. (1981) *A second course in stochastic processes*, Academy press, New York.

[64] Kendall M.G. (1938) *A new measure of rank correlation*, Biometrika 30, pp. 81-93.

[65] Kohn A.F. (1989) *Dendritic transformations on random synaptic inputs as measured from a neuron's spike train: modelling and simulation*, IEEE Trans. Biomed. Eng. 36, 44-54.

[66] Lachal A. (1991) *On the first passage time for integrated Brownian motion* (in French), Ann. I. H. P., Sect. B 27(3), 385-405.

[67] Lansky P. and Rospars J.P. (1993) *Stochastic model neuron without resetting of dendritic potential. Application to the olfactory system*, Biol. Cybern. 69, 283-294.

[68] Lansky P. and Rospars J.P. (1995) *Ornstein-Uhlenbeck model neuron revisited*, Biol. Cybern. 72, 397-406.

[69] Lansky P. and Rodriguez R. (1999) *Two-compartment stochastic model of a neuron*, Phys. D. 132, 267-286.

[70] Lansky P. and Ditlevsen S. (2008) *A review of the methods for signal estimation in stochastic diffusion leaky integrate-and-fire neuronal models*, Biol. Cybern. 99, 253-262.

[71] Lapique L. (1907) *Reserches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization*, J. Physiol. Pathol. Gen. 9, 620-635.

[72] Lawler G.F. (2006) *Introduction to stochastic processes*, Chapman & Hall, Boca Raton, FL.

[73] Lebedev M.A. and Nelson R.J. (1996) *High-frequency vibratory sensitive neurons in monkey primary somatosensory cortex: entrained and non-entrained responses to vibration during the performance of vibratory-cued hand movements*, Exp. Brain Res, 111, 313-325.

[74] Levine M.W. (1980) *Firing rate of a retinal neuron are not predictable from interspike interval statistics*, Biophysical journal, 30, 9-25.

[75] Levy G., Gibaldi M. and Jusko W.J. (1969) *Multi-compartment pharmacokinetic models and pharmacologic effects*, J. Pharm. Sci. 58, 422424.

[76] Linz P. (1985) *Analytical and numerical methods for Volterra equations*, SIAM, Philadelphia.

[77] Majewska A.K., Newton J.R. and Sur M. (2006) *Remodeling of synaptic structure in sensory cortical areas in vivo*, J Neuroscience, 26(11), 3021-3029.

[78] Mood A.M. (1950) *Introduction to the theory of statistics*, McGraw-Hill, New York.

[79] Morozov Y., Khalilov I., Ben-Ari Y. and Represa A. (2002) *Correlative fluorescence and electron microscopy of biocytin-filled neurons with a preservation of the postsynaptic ultrastructure*, Journal of neuroscience methods, 117(1), 81-85.

[80] Mountcastle V.B. (1957) *Modality and topographic properties of single neurons of cat's somatosensory cortex*, J. Neurophysiol., 20, 408-434.

[81] Mulholland R.J. and Keener M.S. (1974) *Analysis of linear compartment models for ecosystems*, Journal of Theoretical Biology, Volume 44, Issue 1, 105-116.

[82] Nawrot M.P., Boucsein C., Rodriguez-Molina V., Aertsen A., Grn S. and Rotter S. (2007) Serial interval statistics of spontaneous activity in cortical neurons "in vivo" and "in vitro", Neurocomputing 70(10), 1717-1722.

[83] Neiman A. and Russell D.F. (2004) *Two distinct types of oscillators in electroreceptors of paddlefish*, Journal of Neurophysiology, 92, 492-509.

[84] Nelsen R.B. (1999) *An introduction to copulas*, Springer, New York.

[85] Oliva A.A. Jr., Lam T.T. and Swann J.W. (2002) *Distally directed dendrotoxicity induced by kainic acid in hippocampal interneurons of green fluorescent protein-expressing transgenic mice*, J Neuroscience, 22(18), 8052-8062.

[86] Ould-Said E. (1997) *A Note on Ergodic Processes Prediction via Estimation of the Conditional Mode Function*, Scandinavian Journal of Statistics 24, 231-239.

[87] Overdijk J., Uylings H.B.M., Kuypers K. and Kamstra A.W. (1978) *An economical, semi-automatic system for measuring cellular tree structures in three dimensions, with special emphasis on Golgi-impregnated neurons*, J Microsc 114, 271284.

[88] Nobile A.G., Ricciardi L.M. and Sacerdote L. (1985) *Exponential trends of first passage time densities for a class of diffusion processes with steady-state distribution*, J. Appl. Prob. 22, 611-618.

[89] Panfilo G., Tavella P and Zucca C. (2004) *Stochastic Processes for modelling and evaluating atomic clock behaviour*, Advanced Mathematical and Computational Tools in Metrology VI, P. Ciarlini, M. G. Cox, F. Pavese ed., World Scientific Publishing.

[90] Parzen E. (1962) *On the estimation of a probability density and mode*, Ann. Math. Statist. 33, 1065-1076.

[91] Peskir G. (2002) *Limit at zero of the Brownian first-passage density*, Probab. Theory Related Fields. 124, 100-111.

[92] Prakasa Rao B.L.S. and Van Ryzin J. (1985) *Asymptotic theory for two estimators of the generalized failure rate*, Statistical Theory and Data Analysis, ed. K. Matusita, North Holland, Amsterdam, 547-563.

[93] Redner S. (2001) *A guide to first-passage processes*, Cambridge University Press.

[94] Reiss R.D. (1981) *Nonparametric estimation of smooth distribution functions*, Scand. J. Statist., 405 (8), 116119.

[95] Ricciardi L.M. (1977) *Diffusion processes and related topics in: Lecture Notes in Biomathematics*, Springer Verlag, Berlin.

[96] Ricciardi L.M. and Sacerdote L. (1979) *The Ornstein-Uhlenbeck process as a model for neuronal activity*, Biol.Cyb. 35, 1-9.

[97] Ricciardi L.M., Sacerdote L. and Sato S. (1984) *On an Integral Equation for First-Passage-Time Probability Densities*, Journal of Applied Probability, Vol. 21, No. 2, 302-314.

[98] Ricciardi L.M. and Sato S. (1988) *First-passage-time density and moments of the Ornstein-Uhlenbeck process*, Journal of Applied Probability, 43-57.

[99] Ricciardi L.M. and Sato S. (1990) *Diffusion processes and first-passage-time problems* in: L.M. Ricciardi (Ed.), Lectures in Applied Mathematics and Informatics, Manchester Univ. Press., Manchester.

[100] Rice J. and Rosenblatt M. (1976) *Estimation of the log survivor function and hazard function*, Sankhya Ser. A, 38, 60-78.

[101] Rieke F., Warland D. and Steveninck R. (1997) *Spikes: Exploring the neural code*, Cambridge, MIT Press, MA.

[102] Rogers L.C.G. and Williams D. (1994) *Diffusions, Markov processes, and martingales. Vol. 1: Foundations*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester.

[103] Román P., Serrano J.J. and Torres F. (2008) *First-passage-time location function: Application to determine first-passage-time densities in diffusion processes*, Computational Statistics & Data Analysis, Volume 52, Issue 8, 4132-4146.

[104] Rosenblatt M. (1956) *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. 27, 832-835.

[105] Rosenblatt M. (1971) *Curve estimates*, Ann. Math. Statist. 42(6), 1815-1842.

[106] Rospars J.P., Lansky P., Vaillant J., Duchamp-Viret P. and Duchamp A. (1994) *Spontaneous activity of first-and second-order neurons in the olfactory system*, Brain Res. 662, 31-44.

[107] Ross S.M. (1996) *Stochastic processes*, John Wiley and Sons, New York.

[108] Sacchetto L. (2009) *Tempo di primo passaggio del moto browniano integrato: aspetti analitici, numerici e simulativi.* Master degree thesis in Mathematics, University of Torino.

[109] Sacerdote L. (1988) *Asymptotic behaviour of Ornstein-Uhlenbeck first-passage-time density through boundaries*, Applied Stochastic Models and Data Analysis. 6, 53-57.

[110] Sacerdote L. and Tomassetti F. (1996) *On Evaluations and Asymptotic Approximations of First-Passage-Time Probabilities*, Advances in Applied Probability, Vol. 28, No. 1, 270-284.

[111] Sacerdote L. and Giraudo M.T. (2013) *Stochastic Integrate and Fire Models: a Review on Mathematical Methods and their Applications*, In Bachar, Batzel and Ditlevsen (Eds.), Stochastic Biomathematical Models with Applications to Neuronal Modeling (2058), 99-142.

[112] Sandberg I. (1978) *On the mathematical foundations of compartmental analysis in biology, medicine, and ecology*, IEEE Trans. Circuits and Systems, 25(5), 273-279.

[113] Sato S. (1977) *Evaluation of the first passage time probability to a square root boundary for the Weiner process*, J. Appl. Prob. 14, 850-856.

[114] Shumway R.H. and Stoffer D.S. (2000) *Time series analysis and its applications*, Springer.

[115] Siegert A.J.F. (1951) *On the first passage time probability problem*, Physical Rev. (2) 81, 617623.

[116] Silverman B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[117] Sklar A. (1959) *Functions de repartition a n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris 8, 229-231.

[118] Sklar A. (1973) *Random variables, joint distribution functions and copulas*, Kybernetika 9(6), 449-460.

[119] Smith P.L. (2010) *From Poisson shot noise to the integrated Ornstein Uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time*, J. Math. Psych. 54, 266-283.

[120] Snow P.J., Rose P.K. and Brown A.G. (1976) *Tracing axons and axon collaterals of spinal neurons using intracellular injection of horseradish peroxidase*, Science 191(4224), 312-313.

[121] Stein R.B. (1965) *A theoretical analysis of neuronal variability*, Biophys. Journal 5, 385-386.

[122] Stevens C. and Zador A. (1998) *Novel integrate-and-fire-like model of repetitive firing in cortical neurons*, In Proc. of the 5th Joint Symposium on Neural Computation.

[123] Trachtenberg J.T., Chen B.E., Knott G.W., Feng G., Sanes J.R., Welker E. and Svoboda K. (2002) *Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex*, Nature, 420(6917), 788-794.

[124] Tran L.T. (1989) *The L_1 convergence of kernel density estimates under dependence*, Canad. J. Statist. 17, 197-208.

[125] Tuckwell Henry C. (1988) *Introduction to theoretical neurobiology, Volume 1: linear cable theory and dendritic structure*, Cambridge University Press.

[126] Tuckwell Henry C. (1988) *Introduction to theoretical neurobiology, Volume 2: Nonlinear and stochastic theories*, Cambridge University Press.

[127] Varadhan S.S. (2007) *Stochastic processes*, Courant Lecture Notes in Mathematics, Vol. 16, Courant Institute of Mathematical Sciences (NY), American Mathematical Society, Providence.

[128] Von Mises R. (1931) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik*, Leipzig and Wien, Franz Deuticke.

[129] Wang L. and Pötzelberger K. (2007) *Crossing Probabilities for Diffusion Processes with Piecewise Continuous Boundaries*, K. Methodology and Computing in Applied Probability, Volume 9, Number 1, 21-40.

[130] Watson G.S. and Leadbetter M.R. (1964) *Hazard analysis I*, Biometrika, 51, 175-184.

[131] Watson G.S. and Leadbetter M.R. (1964) *Hazard analysis II*, Sankhya Ser. A, 26, 110-116.