Multivariate Multiscale Analysis of Neural Spike Trains

by

Reza Ramezan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2013

© Reza Ramezan 2013

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This dissertation introduces new methodologies for the analysis of neural spike trains. Biological properties of the nervous system, and how they are reflected in neural data, can motivate specific analytic tools. Some of these biological aspects motivate multiscale frameworks, which allow for simultaneous modelling of the local and global behaviour of neurons. Chapter 1 provides the preliminary background on the biology of the nervous system and details the concept of information and randomness in the analysis of the neural spike trains. It also provides the reader with a thorough literature review on the current statistical models in the analysis of neural spike trains. The material presented in the next six chapters (2-7) have been the focus of three papers, which have either already been published or are being prepared for publication.

It is demonstrated in Chapters 2 and 3 that the multiscale complexity penalized likelihood method, introduced in Kolaczyk and Nowak (2004), is a powerful model in the simultaneous modelling of spike trains with biological properties from different time scales. To detect the periodic spiking activities of neurons, two periodic models from the literature, Bickel et al. (2007, 2008); Shao and Lii (2011), were combined and modified in a multiscale penalized likelihood model. The contributions of these chapters are (1) introducing a powerful visualization tool, inter-spike interval (ISI) plot, (2) combining the multiscale method of Kolaczyk and Nowak (2004) with the periodic models of Bickel et al. (2007, 2008) and Shao and Lii (2011), to introduce the so-called additive and multiplicative models for the intensity function of neural spike trains and introducing a cross-validation scheme to estimate their tuning parameters, (3) providing the numerical bootstrap confidence bands for the multiscale estimate of the intensity function, and (4) studying the effect of time-scale on the statistical properties of spike counts.

Motivated by neural integration phenomena, as well as the adjustments for the neural refractory period, Chapters 4 and 5 study the Skellam process and introduce the Skellam Process with Resetting (SPR). Introducing SPR and its application in the analysis of neural spike trains is one of the major contributions of this dissertation. This stochastic process is biologically plausible, and unlike the Poisson process, it does not suffer from limited dependency structure. It also has multivariate generalizations for the simultaneous analysis of multiple spike trains. A computationally efficient recursive algorithm for the estimation of the parameters of SPR is introduced in Chapter 5. Except for the literature review at the beginning of Chapter 4, the rest of the mate-

rial within these two chapters is original. The specific contributions of Chapters 4 and 5 are (1) introducing the Skellam Process with Resetting as a statistical tool to analyze neural spike trains and studying its properties, including all theorems and lemmas provided in Chapter 4, (2) the two fairly standard definitions of the Skellam process (homogeneous and inhomogeneous) and the proof of their equivalency, (3) deriving the likelihood function based on the observable data (spike trains) and developing a computationally efficient recursive algorithm for parameter estimation, and (4) studying the effect of time scales on the SPR model.

The challenging problem of multivariate analysis of the neural spike trains is addressed in Chapter 6. As far as we know, the multivariate models which are available in the literature suffer from limited dependency structures. In particular, modelling negative correlation among spike trains is a challenging problem. To address this issue, the multivariate Skellam distribution, as well as the multivariate Skellam process, which both have flexible dependency structures, are developed. Chapter 5 also introduces a multivariate version of Skellam Process with Resetting (MSPR), and a so-called profile-moment likelihood estimation of its parameters. This chapter generalizes the results of Chapter 4 and 5, and therefore, except for the brief literature review provided at the beginning of the chapter, the remainder of the material is original work. In particular, the contributions of this chapter are (1) introducing multivariate Skellam distribution, (2) introducing two definitions of the Multivariate Skellam process in both homogeneous and inhomogeneous cases and proving their equivalence, (3) introducing Multivariate Skellam Process with Resetting (MSPR) to simultaneously model spike trains from an ensemble of neurons, and (4) utilizing the so-called profile-moment likelihood method to compute estimates of the parameters of MSPR.

The discussion of the developed methodologies as well as the "next steps" are outlined in Chapter 7.

Acknowledgements

I wish to thank my supervisors Dr. Paul Marriott and Dr. Shoja'eddin Chenouri for their inspiration, insightful guidance, and generous financial support. I also extend sincere thanks to Dr. Matthijs van der Meer from of the Department of Biology and Dr. Britt Anderson of the Department of Psychology, both of the University of Waterloo, for sharing their insight and valuable comments about the biology of the nervous system. I particularly wish to thank Dr. Matthijs van der Meer for providing me with several spike train datasets and for the many meetings during which he introduced a variety of neuroscience research topics to me. Special thanks are also due to Dr. Lawrence Sincich of the University of Alabama for providing the retinogeniculate synapse data and for his clarifications on the data during our discussions about RGC and LGN neurons.

I am thankful to Dr. Christopher Small, Dr. Cecilia Cotton, Dr. Matthijs van der Meer of the University of Waterloo and Dr. Michael Lavine of the University of Massachusetts Amherst for siting on the thesis committee, and for their insightful comments on earlier drafts of this dissertation.

I would like to thank my very good friends, Vahed Maroufy and Yasaman Hosseinkashi. Their support both research-based and social, not to mention their friendship, made my PhD program years ones I will always cherish in fond memory. I am also thankful to Amirhossein Vakili and Adrian Waddale for sharing their comments and helping with a few parts of the codes developed for this dissertation.

I am truly grateful to my parents and to my siblings. Special thanks to my loving parents Mostafa Ramezan and Nasrin Vaziri for their constant encouragement and support in all stages of my studies. I would like to warmly thank my brother Babak Ramezan and my sister Layla Ramezan whose emotional support and assistance have been beyond what one could expect.

It is impossible to overstate my loving gratitude to my wife, Gwynedd, for her constant love, care and support during my PhD studies. Her patience and encouragement eased the preparation of this dissertation. I owe Gwynedd and her father, Dr. John. A. MacLeod, special thanks for proofreading this dissertation. Needless to say, I am solely responsible for all remaining mistakes.

To my caring parents, Nasrin and Mostafa

To my better half, the love of my life, and my best friend, Gwynedd

Contents

List of Figures **xii**

List of Tables **xiv**

1 Preliminaries and Literature Review **1**

1.1 Introduction 1
1.2 Neuroanatomy and Neurophysiology 2
1.2.1 Early research 2
1.2.2 Neurons and supporting cells 3
1.3 Neural communication 4
1.3.1 The generation of the action potential 5
1.3.2 Synaptic transmission 8
1.4 Spatial and temporal scales of the brain 11
1.5 Randomness and information coding 12
1.5.1 Sources of randomness 13
1.5.2 Neural information coding 14
1.6 Statistical analysis of neural spike trains 17
1.6.1 Integrate-and-Fire model (IF) 18
1.6.2 Point processes, histogram-based, and filtering models 19
1.6.3 Filtering/Smoothing models 21
1.6.4 Renewal processes 22
1.6.5 Likelihood-based inference 24
1.6.6 Nonparametric and Bayesian inference 25
1.7 Neural data collection 26
1.7.1 Recordings from populations of neurons 27
1.7.2 Single cell recordings 28
1.8 Discussion 30
1.9 Online data resources 31

2 Multiscale Analysis of Neural Spike Trains **32**
2.1 Introduction . . . 32
2.2 Statement of the problem . . . 32
2.3 Data visualization: interspike interval (ISI) plots . . . 34
2.4 Multiscale Modelling . . . 38
2.4.1 Multiresolution Analysis for Likelihoods . . . 38
2.4.2 A multiresolution probability model for counts . . . 42
2.4.3 Multiscale penalized likelihood . . . 43
2.4.4 Choosing tuning parameters via cross-validation . . . 43
2.5 Modelling periodicity and brain rhythms . . . 46
2.6 Discussion . . . 50
2.7 Cross-validation and likelihood derivations . . . 53
2.7.1 Derivation of the cross-validation method . . . 53
2.7.2 Loglikelihood derivation for the multiplicative model . . . 54
2.7.3 Loglikelihood derivation for the additive model . . . 55

3 Data Analysis Using Multiscale Poisson Models **57**
3.1 Introduction . . . 57
3.2 Retinogeniculate synapse data analysis . . . 57
3.3 Discussion . . . 68
3.4 Supplementary plots and frequency values for other neurons . . . 69
3.4.1 Summary of the multiscale fits for other neurons . . . 69

4 Univariate Skellam Process With Resetting **73**
4.1 Introduction . . . 73
4.1.1 Neural inhibition . . . 73
4.2 Skellam distribution and Skellam process . . . 74
4.2.1 Homogeneous Skellam process . . . 75
4.2.2 Inhomogeneous Skellam process . . . 77
4.3 Skellam process with resetting (SPR) . . . 80
4.4 Skellam process and Markov chain . . . 82
4.4.1 Over dispersion . . . 84
4.5 The effect of time-scale . . . 86
4.5.1 Interspike interval distribution . . . 86
4.5.2 Connection with Brownian motion . . . 87
4.5.3 Large time-scales . . . 88
4.6 Spike count and ISI variability . . . 90
4.7 Discussion . . . 92

4.8 Proofs of the theorems and some preliminary results on the mean-variance relationship . . . 93
4.8.1 Proof of Theorem 4.1 . . . 93
4.8.2 Proof of Theorem 4.2 . . . 95
4.8.3 Proof of Theorem 4.3 . . . 95
4.8.4 Proof of Equation (4.5.15) . . . 97
4.8.5 Some preliminary results on the mean-variance relationship in SPR 98

5 Parameter Estimation in Univariate SPR 104
5.1 Introduction . . . 104
5.2 Likelihood function . . . 104
5.2.1 A computationally tractable method for parameter estimation . 109
5.2.2 Two intervals: before the first and after the last spike . . . 110
5.3 Data analysis . . . 112
5.3.1 Simulation study . . . 112
5.3.2 Real data analysis . . . 114
5.4 Multiscale estimation . . . 115
5.4.1 Tuning the parameters via cross-validation . . . 116
5.4.2 Simulation study . . . 117
5.4.3 Real data analysis . . . 118
5.5 Discussion . . . 121

6 Multivariate Skellam Process With Resetting 124
6.1 Introduction . . . 124
6.2 Conventional multivariate models . . . 124
6.3 Multivariate Skellam Distribution . . . 126
6.4 Multivariate Skellam process . . . 130
6.4.1 Bivariate homogeneous Skellam process . . . 130
6.4.2 Dimensionality of the parameter space . . . 134
6.4.3 Inhomogeneous multivariate Skellam process . . . 134
6.5 Multivariate Skellam process with resetting (MSPR) . . . 136
6.6 Parameter estimation . . . 137
6.6.1 Method of moments and likelihood . . . 137
6.7 Data analysis . . . 139
6.7.1 Simulation study . . . 139
6.7.2 Real data study . . . 140
6.8 Discussion . . . 141
6.9 Proofs of the theorems and a parameter estimation algorithm . . . 142

6.9.1 Proof of Theorem 6.1 142
6.9.2 Proof of Theorem 6.2 149
6.9.3 The recursive parameter estimation algorithm 151

7 Discussion and Future Work **157**
7.1 Introduction 157
7.2 Multiscale analysis within Poisson framework 157
7.3 Univariate Skellam process with resetting (SPR) 158
7.4 Multivariate Skellam process with resetting (MSPR) 160
7.5 Derivation of spike count distribution 163

Bibliography **165**

List of Figures

1.1 The main parts of a neuron . . . 3
1.2 The six stages of an action potential generation process . . . 6
1.3 Synapse: the information transmission structure . . . 8
1.4 The relationship between the strength of a stimulus and the firing rate 10
1.5 Neural integration . . . 11
1.6 Different spatial scales of the brain . . . 12
1.7 Sample PSTH . . . 21
1.8 Different ISI Distributions . . . 23
1.9 The recording unit housing microelectrodes used in data collection . . . 29
1.10 The spatiotemporal resolutions of neural activity recording methods . . . 30

2.1 Sample ISI plots based on real data . . . 36
2.2 ISI plot of the real data vs. that of the simulated homogeneous Poisson process . . . 37
2.3 Dyadic and non-dyadic recursive partitioning of the interval $[0, 1)$ . . . 40
2.4 The performance of the cross validation method: unbiasedness and reconstruction power . . . 45
2.5 Multiscale intensity function of an inhomogeneous Poisson process, $\theta(t) = c(t)\Big(2 + \cos(8\sqrt{2}\pi t)\Big)$ . . . 48
2.6 Comparison between the true and multiscale estimate of the intensity function . . . 51
2.7 The windowing effect . . . 52

3.1 The raster plot of the first 25 replications of the data from ON-center Parvocellular cells . . . 59
3.2 The average of the multiscale fit of the intensity function (inhomogeneous Poisson) for the ON-center Parvocellular cells data . . . 62
3.3 The effect of averaging over trials on cross-validation . . . 62
3.4 The multiscale fit (inhomogeneous Poisson) for an individual trial of the ON-center Parvocellular cells data . . . 63

3.5 The Fourier transform of the stimulus signal 64
3.6 The additive, multiplicative and BARS fits for the ON-centre Parvocellular cells data 66
3.7 Kolmogorov-Smirnov goodness-of-fit test for BARS, the additive, and the multiplicative models 67
3.8 The multiscale fit (Poisson) for the Parvocellular cells 70
3.9 The multiscale fit (Poisson) based on Marvocellular cells data 71

4.1 Examples of homogeneous Skellam process 77
4.2 Examples of inhomogeneous Skellam process 79
4.3 A toy example plotting the trajectory of a Skellam Process with Resetting 83
4.4 Mean-variance relationship in the equilibrium distribution of Skellam Process with Resetting 85
4.5 Parameter space under which $\Pr(N_T \leq 0|\lambda_1, \lambda_2) < \epsilon$ 89
4.6 Distribution of spike counts in large time scales 89
4.7 Mean-variance relationship in the Skellam process with resetting 90
4.8 Interspike interval quantiles: Poisson vs SPR interspike intervals 91

5.1 Possible unobserved paths for given observable data 106
5.2 Simulated data from a homogeneous SPR 113
5.3 Scatter plot of the parameter estimates based on simulated data 114
5.4 Raster plot and parameter estimates from retinogeniculate synapse data 115
5.5 Simulated data from a inhomogeneous SPR 118
5.6 The average multiscale estimates based on simulated data 119
5.7 The average multiscale estimates based on retinogeniculate synapse data 120
5.8 Raster plots for retinogeniculate synapse data (simulated vs real data) . 121
5.9 Mean spiking activity and the stimulus signal 122

6.1 The probability mass function of a bivariate Skellam random vector . . 131
6.2 The effect of correlation on probabilities in bivariate Skellam process . 133
6.3 The raster plot of the first 25 trials of the on-center parvocellular recordings 141
6.4 An example of bivariate spike trains 154

7.1 An under-dispersed model 159
7.2 Simulated Skellam process wth resetting with parameter values $(\lambda_1, \lambda_2) = (10, 9)$ 161
7.3 Spike counts and conditional Skellam distribution 162

List of Tables

2.1 Comparison of wavelet and likelihood MRA 42
2.2 Simulation results of model selection criteria for additive and multiplicative models 50

3.1 Cross-validation results for the pair (N^*, λ^*) 60
3.2 Spike count correlation between RGC and LGN at different resolutions 61
3.3 The 5 most repeated frequencies for ON-centre parvocellular cells data across the 129 trials 64
3.4 Parameter estimates for the multiplicative and additive models based on RGC and LGN data 65
3.5 The 5 most repeated frequencies for ON-centre Parvocellular cells across the 129 trials 72
3.6 The 5 most repeated frequencies for ON-centre Magnocellular data across the 129 trials 72
3.7 The 5 most repeated frequencies for OFF-centre Parvocellular cells across the 129 trials 72
3.8 The 5 most repeated frequencies for OFF-centre Magnocellular data across the 129 trials 72

4.1 Moments of the interspike intervals (ISI) based on real data, and simulated Poisson and SPR 92

5.1 The parameter values a_i, α_i, β_i and γ_i in a toy example 108
5.2 Point and interval estimates of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ based on simulated data . . . 113
5.3 Point and interval estimates of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ based on the ON-centre Parvocellular data 116

6.1 Mapping between the moment space and the parameter space for Skellam distribution 138
6.2 Parameter estimates for a simulated homogeneous Skellam process with resetting 139

6.3 Parameter estimates for the homogeneous Skellam process with resetting based on the ON-center Parvocellular data 142

Chapter 1

Preliminaries and Literature Review

1.1 Introduction

The brain, the centre of the nervous system, is one of the most complex organs in human body. Although it constitutes approximately 2% of our total body weight, due to the high metabolism of its nerve cells, the human brain receives about 15% of the cardiac output and consumes 20% of total body oxygen, as well as 25% of total body glucose utilization, Carlson (2007); Magistretti et al. (1995). It regulates our emotions, thoughts, behaviour, perception, motor movements and the function of all organs, Murray (2007); LeDoux (2003); Lipton (2005); Kanai et al. (2011); Knapen et al. (2011); Milpass (2012). Although the literature on brain research is significantly rich, many questions about the human brain remain to be answered. The work of van Hemmen and Sejnowski (2006) discusses some of the open problems in different areas of systems neuroscience such as brain evolution, organization of the cerebral cortex, interaction between neurons, computation in the brain and organization of cognitive systems. Being able to fully cure patients suffering from brain-related illnesses such as Alzheimer's and dementia, or even to build an artificial brain, also motivate ongoing neuroscience research. Two hundred years ago, heart transplants and artificial hearts were science fiction, whereas many lives are being saved by heart transplants and by means of artificial hearts today. A similar transition could happen with brain research. Many attempts have already been made to simulate the human brain. de Garis et al. (2010); Eliasmith (2013) provide interesting examples of large-scale simulations of the brain. These examples include, but are not limited to, Markram's Blue Brain project, Markram (2006), which simulates a cortical column at the level of ion channel details and sizes in at 1 million neurons. Two other projects are Modha's Cognitive Com-

putation Project, Ananthanarayanan and Modha (2007) ($\sim$ 1 billion neurons) and Izhikevich's large-scale cortical simulations, Izhikevich and Edelman (2008) ($\sim$ 100 billion neurons), which are both much larger in terms of the number of neurons, however, since random connectivity of neurons has been employed in these simulations, the output is not easy to interpret and does not correspond to any particular function or behaviour. Eliasmith et al. (2012); Eliasmith (2013) have introduced the Semantic Pointer Architecture Unified Network (Spaun), which is a model of about 2.5 million neurons with image sequences as inputs whose output is a motor behaviour. Spaun is currently the largest functional model of the brain. These research studies would suggest that state-of-the-art research is certainly moving toward building an artificial human brain. In order to proceed, let us introduce some nervous system terminology starting with neuroanatomy and neurophysiology.

Neuroanatomy is the study of the structure of the nervous system while *neurophysiology* is the study of its functions. These two areas include the study of the structure of the neurons, their communication and their underlying electrochemical processes, Carlson (2007); Blum and Rutkove (2007). The purpose of this chapter is to familiarize the reader with these two areas and to provide a literature review of related research. This literature review also discusses the important problem of information coding in the brain. We will discuss the sources of randomness in the neural data, as well as the common statistical models in the analysis of neural data. Some details about data collection techniques are also presented.

1.2 Neuroanatomy and Neurophysiology

The nervous system consists of all nerve cells, the brain and the spinal cord. It is usually divided into two parts: the central nervous system (CNS), and the peripheral nervous system (PNS). The brain and the spinal cord together form the CNS and the remaining part of the nervous system, which includes the nerves attached to the CNS, forms the PNS.

1.2.1 Early research

The discovery and development of the foundations of modern neuroscience go back to the 19$^{\text{th}}$ century. According to Pearce (2001), in 1848, Emil du Bois-Reymound discovered the neural impulse (action potential), whose conduction velocity was first measured by his friend Hermann von Helmholtz in 1850. By the end of the 19$^{\text{th}}$ century, Santiago Ramón y Cajal introduced the later-called "neuron theory" and revealed

the shapes of the nerve cells by using a stain developed by Camillo Golgi, López-Muñoz et al. (2006). The two scientists shared the 1906 Nobel Prize in Physiology and Medicine for their work on the structure of the nervous system. However, it was the mid 19^{th} and 20^{th} century anatomist Heinrich Wilhelm Gottfried von Waldeyer-Hartz who first used the term *neuron* in reference to nerve cells, Ramón y Cajal (1954). In a 1926 monograph entitled "The Interactive Action of the Nervous System," Charles Sherrington explained that the impulse in the chemical synapse (small gap between two adjacent neurons) flow in one direction only, Sherrington (1926). He developed a theory, which declares that the nervous system coordinates different parts of the body and states that interactive actions of the nervous system are observed as reflexes in the body. Charles Sherrington and Edgar Adrian shared the 1932 Nobel prize in Physiology and Medicine for their discoveries regarding the functions of neurons.

1.2.2 Neurons and supporting cells

Neurons are special cells which are both information-processing and information-transmitting units of the nervous system. They consist of a *soma* (or cell body), *axon*, *dendrites* and *terminal buttons*. The soma provides the energy for the cell and contains genetic information. Two structures connect to the soma: the axon and the dendrites. Dendrites are tree-shaped structures which *usually* receive information from other neurons and take it back to the soma. The axon, which is a long and thin tube, takes information from the soma and passes it to other neurons. Terminal buttons, found at the end of the twigs of an axon, facilitate the transmission of information from the sending cell to the receiving cell, Carlson (2007). Figure 1.1 shows a neuron.

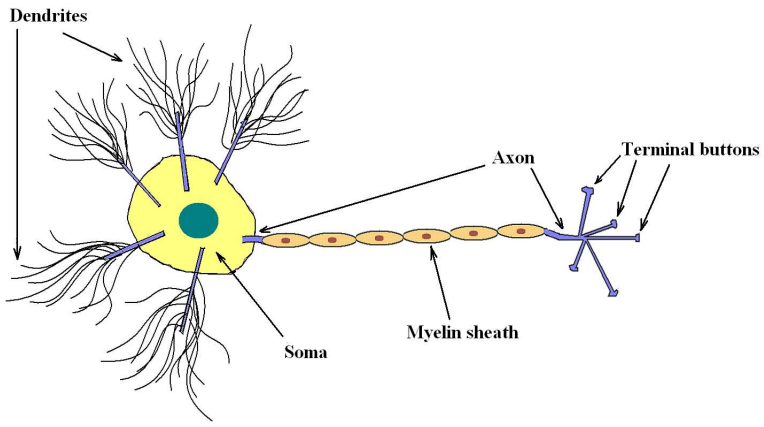


Figure 1.1: *The main parts of a neuron.*

The shape and the function of a neuron depends on the job it performs. As a result, neurons are divided into three categories according to their appearance. *Multipolar neurons* are the most common neurons in the CNS with one axon and many dendrites attached to the soma. *Bipolar neurons* have one axon and one dendrite attached to the soma and are primarily found in sensory systems such as vision and hearing. *Unipolar neurons* have only one axon attached to the soma and are found in the somatosensory systems such as touch and pain.

In terms of function, neurons are also divided into three categories. *Sensory neurons* receive information and stimuli from the internal or external environment, translate them to electrochemical signals, and send these signals to the central nervous system. They form a large part of the peripheral nervous system. *Motor neurons* located within the central nervous system control the contraction of muscles or secretion from glands. Finally are the *interneurons*, which are entirely located within the central nervous system, receive information from sensory neurons and pass it along to motor neurons. They are also involved in learning, memory, perceiving, deciding and controlling complex behaviour. For more details on the shape and function of neurons see Nicholls et al. (2012); Carlson (2007).

Neurons are not the only cells found in the nervous system. About a half of the CNS is formed by other cells called *supporting cells*. These cells supply neurons with nutrients and oxygen and help them perform their jobs. In the CNS, *glial cells* are the most important supporting cell, in that they work as protective shields, glue the neurons together and hold them in place, supply the nutrients, and literally "digest" dead neurons and clean debris. Different types of glial cells (Astrocyte, Oligodendrocyte and Microglia) are responsible for these different tasks. As an example, Oligodendrocytes support axons and produce myelin. In the PNS, *Schwann cells* are the most important supporting cells and are Oligodendrocytes counterparts. For more readings on supporting cells refer to Kandel et al. (2000).

1.3 Neural communication

A neural message, which is an electrochemical wave, travels along the axon of a neuron to the terminal buttons. As can be seen in Figure 1.1, a bead-shaped structure (the myelin sheath) cover the axon. The function of the myelin sheath is to speed up this transition and insulate the signal, preventing it from spreading between consecutive axons, (Carlson, 2007, p.38). This signal is then transmitted to another neuron in

a structure formed by the terminal buttons of the sending neuron and part of the dendritic or somatic membrane of the receiving neuron. This structure is called a *synapse*. The communication process consists of two subprocesses:

1. signal generation in the sending neuron and its conduction to the terminal buttons of this neuron;

2. transmission of the message at the synapse.

Subsections 1.3.1 and 1.3.2 discuss these two subprocesses.

1.3.1 The generation of the action potential

The membranes of all cells (including neurons) are electrically charged, i.e., the potential of the inside of a cell, *the intracellular potential*, is different from the potential of the outside of the cell– *the extracellular potential*. Neurons utilize this cross-membrane electrical charge to generate and conduct neural messages. A neural message, which is called an *action potential*, is an electrochemical wave caused by changes in the concentration of the positively and negatively charged ions in the extracellular and the intracellular fluid, Kandel et al. (2000). The cross membrane potential of a neuron, which is called the *resting potential*, is approximately -70mV. The *ion channels*, some of which are voltage-dependent gates, control the dynamics of the inflow and outflow of the positive and negative ions which are called *cations* and *anions* respectively. Naming these channels after the ions passing through them, they are called either sodium, potassium, calcium or chloride channels, Hille (2001).

To understand what causes the action potential, two forces should be introduced: *diffusion* and *electrostatic pressure*. According to Carlson, (Carlson, 2007, p.45-47), while diffusion force moves molecules from high concentration areas to low concentration areas, the electrostatic power is the attractive or repulsive power between electrically charged particles. When a neuron is excited by some stimuli or inputs from other neurons, the membrane potential is *depolarized* i.e., the membrane potential is reduced from the resting potential towards zero. If this depolarization meets the *threshold of excitation*, which is about -60mV, the sodium channels open. As a result, both diffusion and the electrostatic forces lead the sodium *cations* (positive ions), inside the cell. Therefore, the inside potential of the cell starts to decrease (note that the inside potential of the membrane is initially negative). Shortly after the opening of the sodium channels, potassium channels also open and diffusion power pushes potassium cations (positive ions) out of the cell. However, because the membrane is more permeable to sodium ions, the influx of the sodium ions dominates the outflow of the potassium ions.

Therefore, the membrane potential increases to reach its peak at +40mV where the sodium channels become *refractory* i.e., the channels become blocked and cannot open again until the membrane potential once more reaches the resting potential. However, the potassium channels are still open and the diffusion power lets these ions escape. Shortly after, the potassium channels are closed and the extra potassium ions outside the cell will diffuse. The small amount of extra potassium outflow *hyperpolarizes* the membrane for a short period as the potassium ions outside the cell diffuse. *Hyperpolarization* is the process during which the membrane potential is increased from the resting potential. The Nobel prize winners Alan Lloyd Hodgkin and Andrew Fielding Huxley introduced a set of differential equations, which model the dynamics of the ion channels, Hodgkin and Huxley (1952). Figure 1.2 summarizes the process of action potential generation.

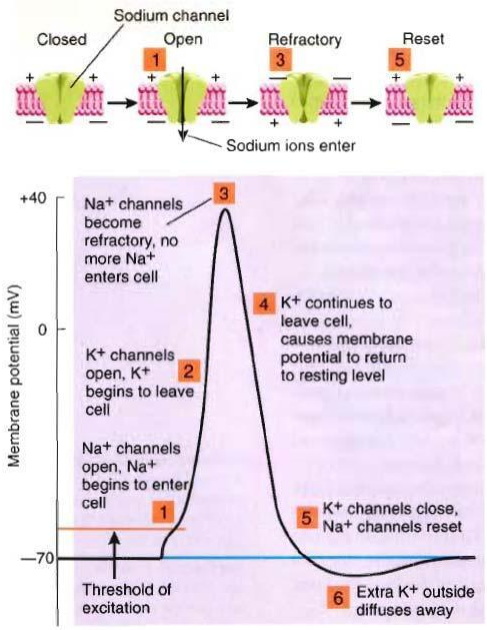


Figure 1.2: *The six stages of an action potential generation process, (Carlson, 2007, p.49). Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.*

Since action potentials tend to be very localized in time, they are called *spikes*. Furthermore, a sequence of spikes generated by a single neuron is called a *spike train*, (Gerstner and Kistler, 2002, p.3). As a result, discussing the randomness in spiking activity and the corresponding statistical issues, the problem is usually formulated in a point process framework. A typical spike train from a neuron looks like $(t_1, t_2, ..., t_n)$ where t_i is the time of the i^{th} spike. Although the shapes and durations of individual

spikes generated by a given neuron can vary, it is generally assumed that the form of the action potential is not as important in information transmission. We will explain the concept of neural information and randomness in Section 1.5.

A second *firing*, or the trigger of an action potential, cannot be initiated immediately after an action potential is generated. A certain amount of time ($\approx 1-3$ milliseconds) must elapse before another spike is fired. This time period is called the *absolute refractory period* and limits the maximum frequency of the neural firing rate to about 1000 spikes per second. The firing rate is defined as the number of spikes per unit of time, (Gerstner and Kistler, 2002, p.14). Another type of refractoriness is the *relative refractory period*, during which the initiation of another action potential is possible but it requires a much higher stimulation. In this dissertation refractory period refers to the absolute refractoriness unless otherwise stated.

One of the inevitable properties of the neural activity is the periodicity, or the rhythms of the brain, Buzsáki (2006), which is observed in different parts of the brain. The biological mechanisms generating these rhythms are of interest to neuroscientists. These rhythms are classified according to their frequencies as follows: δ-rhythm (2-4Hz), Walker (1999), θ-rhythm (4-8Hz), Miller (1991); Malhotra et al. (2012), α-rhythm (8-13Hz), Windhorst and Johansson (1999), β-rhythm (13-30Hz), Lopes da Silva (1991) and γ-rhythm (more than 30Hz), Freeman (1992). Notice that these frequency bands are what we have commonly found in the literature, but they vary from one manuscript to another. Some of these rhythms can be present simultaneously in a given area of the brain. Simultaneous presence of multiple frequencies in a vision experiment on monkeys is reported in Bressler et al. (1993). According to Fischer et al. (2002), θ and γ rhythms characterize the hippocampal activity *in vivo*. For more details on brain rhythms refer to Buzsáki (2006). Chapter 2 introduces a model for oscillatory activities of neural spike trains in an inhomogeneous Poisson process framework (rate code).

After an action potential is triggered, it travels down the axon to the terminal buttons. The amplitude of the action potential is constant (+40mV) from the generation point to the terminal buttons. Moreover, once an action potential is triggered, it will not disappear during transmission. This biological characteristic of neurons is called the *all-or-none law*, whose discovery dates back to Edgar Adrian's work in 1920s. The terminal buttons, which are the main communication ports of neurons, transmit the signal to the other neurons across the *synapse*. The synapse is the minute gap between the terminal buttons of the sending cell and the dendrites of the receiving cell.

1.3.2 Synaptic transmission

The transmission of messages from one neuron to another neuron through a synapse is called *synaptic transmission*. The neuron which conducts the action potential to the synapse is called the *presynaptic neuron*, and the one which receives the signal is called the *postsynaptic neuron*. In the synapse, the space between the membranes of the presynaptic and postsynaptic neurons is called the *synaptic cleft*. Although we have mentioned that a synapse forms between the terminal buttons of the presynaptic neuron and the dendritic membrane of the postsynaptic neuron, synapses can occur in three places: on dendrites (*axodendritic synapse*), on soma (*axosomatic synapse*) and on other axons (*axoaxanic synapse*); however, the axodendritic synapse is the most common type of synapse, Lytton (2002); Weiss et al. (2002). Figure 1.3 shows a synapse.

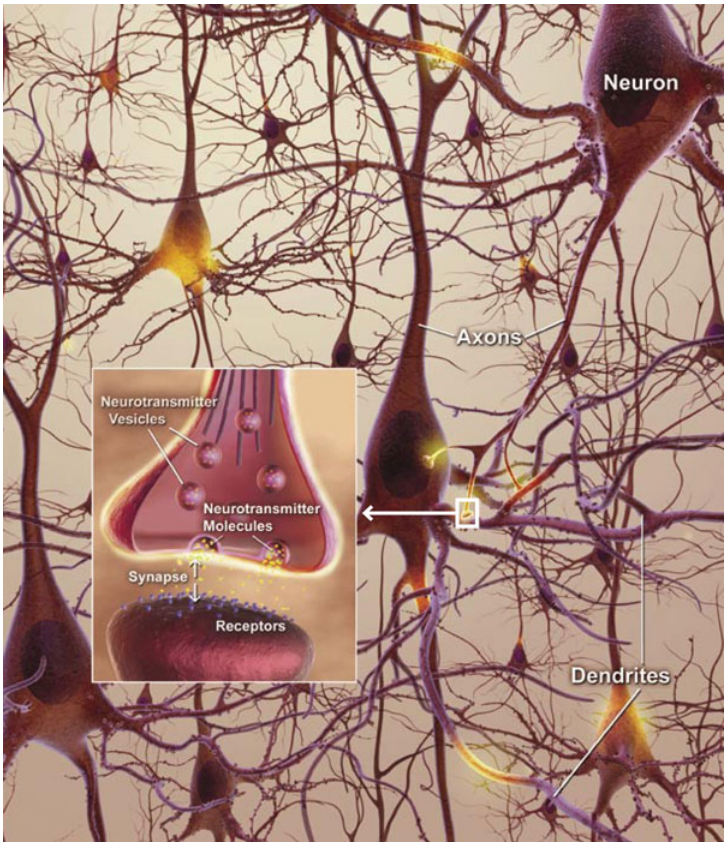


Figure 1.3: *Synapse: the information transmission structure. Reproduced courtesy of the National Institute on Aging, National Institute of Health, USA.*

When the electrochemical wave arrives at the terminal buttons of the presynaptic

neuron, it releases chemicals called *neurotransmitters* in the synaptic cleft. These neurotransmitters then bind with the receptors on the postsynaptic membrane and can depolarize or hyperpolarize the postsynaptic membrane. The alterations in the membrane potential of the postsynaptic neuron caused by the liberation of the neurotransmitters are called *postsynaptic potential*, Purves et al. (2008a). If the released neurotransmitters have an excitatory (inhibitory) effect on the postsynaptic neuron, they may cause an excitatory depolarization (inhibitory hyperpolarization) of the postsynaptic membrane, which is called *excitatory postsynaptic potential (EPSP)* (*inhibitory postsynaptic potential (IPSP)*), Purves et al. (2008a).

Neurons are interconnected by means of synapses through which electrochemical signals pass from one neuron to another. The effect of a stimulus and its strength on the neural activity can be explained through this example. Consider touching two pieces of metal; one hot, one warm. In both cases, some neurons in your somatosensory system are stimulated, however, in the former case the strength of the stimulus is much higher than that of the latter case. The stronger the stimulus, the higher the firing. Therefore, the strength of a stimulus affects the firing rate of a neuron. This property of the neural activity is called *the rate law*. The first research studies on the relationship between a stimulus and neural firing rate are those of Adrian and Zotterman in the mid-1920s, Adrian (1926a,b); Adrian and Zotterman (1926a,b), where they observed a direct relationship between the pressure applied to a patch of skin and the firing rate of peripheral touch receptors. In Chapters 2 and 3 it is assumed that the firing rate codes the information in the brain; however, "information" is yet to be properly defined. Figure 1.4, adapted from (Carlson, 2007, p.51), shows a firing rate representation of the strength of a stimulus.

Another interesting biological fact about neurons is *bursting* activity, Li et al. (2009); Izhikevich (2000); Natarajan (2003). Bursting activity is the sudden and repeated firing of action potentials. Bursting depends on the context, and is more a subjective interpretation of neural activity rather than a well-defined term.[1] As an example, Li et al. (2009) considers the bursting of a rat's cortical neuron as the sudden trigger of 5-15 spikes with no more than 8 ms between these consecutive spikes. However, we have not found any research in the literature which considers a gradual increase in the spiking activity to be bursting. It is important to note that the term "bursting activity" has also been associated in the literature with the activation of T-type calcium channels, Llinás and Steriade (2006). In this dissertation, "bursting" or "burst spiking activity" only refer to a set of consecutive short interspike intervals.

[1] Private discussions with Dr. Britt Anderson of the University of Waterloo.

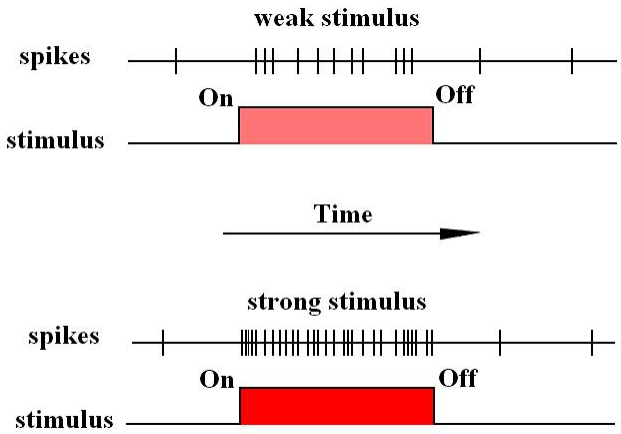


Figure 1.4: *The relationship between the strength of a stimulus and the firing rate.*

A typical neuron forms numerous synapses with other neurons which could have excitatory or inhibitory effects on it. While the excitatory neurons encourage spiking activity in the postsynaptic neuron, increasing the likelihood of spiking, the inhibitory neurons discourage the spiking activity in the postsynaptic neuron. The post synaptic neuron integrates the EPSPs and IPSPs and if the integrated postsynaptic potential crosses the threshold of excitation, an action potential will be triggered. This process is called *neural integration*, Giuliodori and Zuccolilli (2004); Purves et al. (2008a), which is illustrated in Figure 1.5.

As shown previously in Figure 1.4, the rate of the spiking activity depends on the strength of the stimulus. According to (Carlson, 2007, p.62), the relative activity of the excitatory and inhibitory synapses controls the firing rate of the postsynaptic neuron. Furthermore, the increase in the excitatory (inhibitory) synaptic activity increases (decreases) the postsynaptic spiking activity. Notice that *inhibitory/excitatory neurons* and inhibitory/excitatory activities are different concepts. As an example, an inhibitory postsynaptic potential is the result of the activity of an inhibitory neuron. Therefore, the greater the number of presynaptic inhibitory neuron firings, the more excited that neuron is; the more inhibitory effect is generated in the postsynaptic neuron, the less likely the postsynaptic neuron is to fire. This positive or negative interrelation of the spiking activities will be discussed later in Chapters 4 and 6. Throughout this dissertation we will be utilizing statistical techniques and statistical methodologies to analyze neural data. Therefore, it is important to emphasize some of the properties of neural spiking activity which motivate statistical and probabilistic modelling of such data.

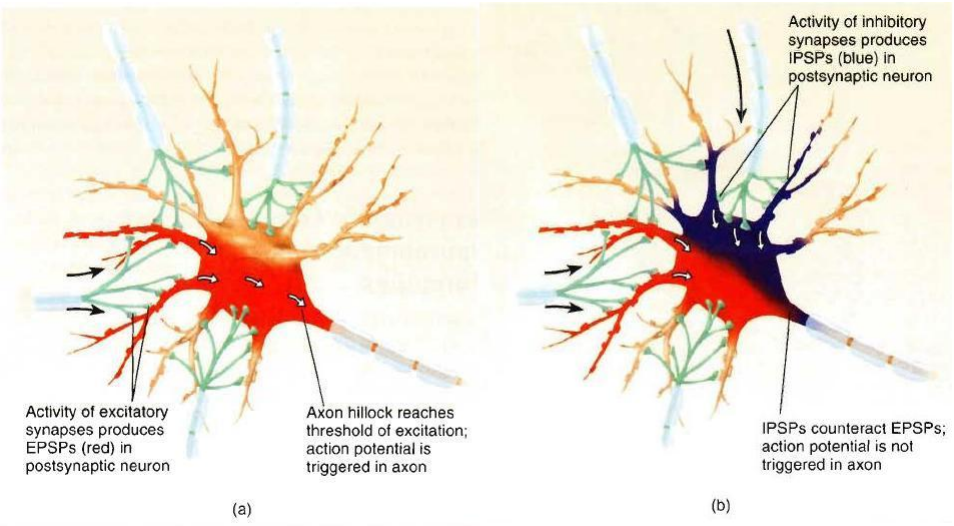


Figure 1.5: *Neural integration, (Carlson, 2007, p.63): (a) If the excitatory synapses dominate the inhibitory ones, the postsynaptic neuron will fire. The EPSPs are shown in red. (b) If the IPSPs produced by the inhibitory synapses (shown in blue) diminish the size of the EPSPs, the postsynaptic neuron will not fire. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.*

1.4 Spatial and temporal scales of the brain

The brain is a multiscale organ in terms of time and space. As far as spatial scale is concerned, while molecules and ion channels are in the scale of nanometres (10^{-9}), systems and the general function of the CNS belong to a much higher scale (10^{0} meters). Figure 1.6 reproduced from Churchland and Sejnowski (1994), illustrates the spatial scales of the CNS.

The main focus of this dissertation is on the neurons or networks. This limits the spatial scale to a specific range shown in Figure 1.6 (10^{-6} to 10^{-3} metres). However, a much wider spectrum in the temporal scale is covered. The time scales of the biological phenomena in the brain are quite variable, from ion channels opening and closing (milliseconds) to development and aging (years). Furthermore, the work of Nelson (2002) shows that the amount of variability in the spike trains depends on the time scale. He showed that the relative refractory effects which are associated with dynamic spike threshold, form the underlying correlation structure of the spike generation process. The two papers Ramezan et al. (2014) and Kass and Ventura (2006) show that (through examples) the spike count correlation (a.k.a. noise correlation) depends on the size of the bins in which the spike are counted. The different time scales of the

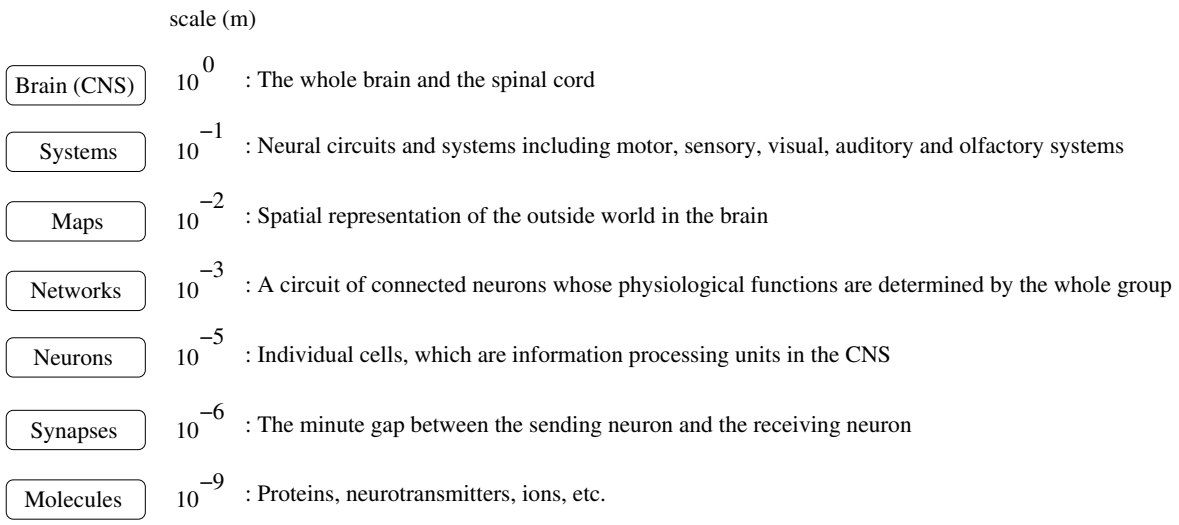


Figure 1.6: *Different spatial scales of the brain.*

biological phenomena, as well as the time scale dependent nature of some of the statistical properties of spike trains, motivate methodologies in the analysis of neural spike trains which take these multiscale behaviours into account. In Chapter 2 we will adapt a multiscale estimation method introduced in Kolaczyk and Nowak (2004, 2005).

1.5 Randomness and information coding

Working with a highly connected network of neurons in a living organ, variability and randomness are inevitable. One type of variability is due to the differences among different subjects in a study. For example, if a particular experiment is performed 50 times on three rhesus monkeys, the differences in the outcomes of the study due to the differences among the three monkeys, is called *subject-to-subject variability*. Another type of variability is observed among the outcomes of repeated trials of an experiment on the same subject. In the example above, the differences in the outcomes of the 50 repeated trials on the same subject is called *trial-to-trial* variability, (Laing and Lord, 2009, ch.6). These types of variabilities are similar to "within" and "between" sum of squares in the Design Of Experiments (DOE). The external environment, the state of the subject, trial-to-trial and subject-to-subject variabilities are usually influential factors in neuroscience research. Furthermore, the very small size and the very large number of neurons makes it hard to measure their electrical activity. Therefore, there is always some measurement error and uncertainty in the extracellular measurements. Due to the very large number of neurons in the brain ($\approx 10^{11}$), in any study only a small number of these neurons are sampled to be observed, resulting in sampling

errors in any study in this research area. Given that the ultimate goal is to crack the neural code, it is important to clarify the sources of randomness and the concept of information in neural activity.

1.5.1 Sources of randomness

According to Moore et al. (1966), an important aspect of the models, which generate spike trains that resemble activity observed experimentally, is the presence of a random component. Some sources of randomness in spike trains are:

1. The fluctuations of the membrane potential (known as *synaptic noise*), as well as the fluctuations in the firing level which are intrinsic to the cell, Moore et al. (1966); Calvin and Stevens (1968); Brunel et al. (2001); Faisal et al. (2008),

2. The random character of the synaptic input arriving at the cell, Moore et al. (1966); Calvin and Stevens (1968),

3. The fundamental stochasticity in the response of neurons to the synaptic input which imposes randomness on the interspike interval, Calvin and Stevens (1968); Moore et al. (1966); Shinomoto et al. (2005),

4. Noisiness of the external sensory stimuli due to their thermodynamic or quantum mechanical nature, Faisal et al. (2008),

5. Measurement errors associated with the recording devices, Jog et al. (2002); Chelaru and Jog (2005).

According to Stein et al. (2005), although the full importance of the variability in the interspike intervals of the individual neurons is yet to be determined, it is definitely not limited to the neural noise. Gerstner and his colleagues, (Gerstner and Kistler, 2002, ch.5), discuss some intrinsic and extrinsic-to-the-cell sources of neural noise. Among the intrinsic sources are the fluctuations in the resistance of the membrane potential. These fluctuations occur because of changes in the temperature and the fact that the number of ion channels in a patch of neural membrane is finite. Synaptic transmission failures (ratio of the presynaptic spikes which don't generate postsynaptic response) is among the extrinsic-to-the-cell sources for neural noise. Apart from the sources of noise related to single cells, the work of Kistler and De Zeeuw (2002) suggests that having a network of connected neurons can also cause irregularity in the spiking activity of neurons. This is known as the network effect. Kostal et al. (2007) discuss some of the properties of spiking randomness as well as an information-theoretic measure of randomness in the spiking activity. They argue that the properties of spiking randomness and variability

are two different matters. While small variability generally implies low randomness, large variability in neural spiking activity does not imply a high level of randomness, i.e., changing the probability model of neural spike trains, one can achieve the same level of randomness with different amounts of variability. For a detailed study on the sources of noise in the nervous system and the effects of noise on the trial-to-trial variability refer to Faisal et al. (2008); Destexhe and Rudolph-Lilith (2012) and the references therein.

1.5.2 Neural information coding

Adrian and Zotterman in a series of studies, Adrian (1926a,b); Adrian and Zotterman (1926a,b), observed a direct relationship between the pressure applied to a patch of skin and the firing rate of peripheral touch receptors. To the best of our knowledge, these are the earliest published works on attempts to understand the relationship between neural activity and events in the outside world, which is called *the neural code*. This is the way the brain receives input, represents the state of the world, and also the way that it sends messages to the body. The neural code is the language of the nervous system and to crack this code, we need to learn the rules of this language. The neural code consists of *encoding* and *decoding* problems. While an encoding problem refers to the way neurons represent a stimulus or behaviour, a decoding problem concerns reproducing the stimulus from neural spike trains, Koyama et al. (2010). Recall the example about the two metal pieces in Section refbetween communication. The encoding problem in touching one of these metal pieces is to understand how the state of the world (texture, temperature, etc.) have been translated into spike trains by the sensory system. On the other hand, the dual decoding problem is to reconstruct the state of the world based on the spike trains generated by neurons.

Now, given a spike train, what can be inferred about the stimulus signal or the state of the sensory world? What aspects of the spike trains are important for the brain to understand the state of the world? To-date no one knows the exact and the complete rules of the encoding-decoding process, but there is a rich literature emphasizing different aspects of neural spike trains, Gerstner and Kistler (2002); Nicolelis (2001); Kostal et al. (2007); Montemurro et al. (2008). The major suggested neural codes are as follows:

1. *Rate Coding* states that the mean or intensity function of the spiking activity summarizes the information of spike trains. Adrian (1926a); Adrian and Zotterman (1926a); Wiesel and Hubel (1959); Adrian (1965); Shadlen and Newsome (1994); Johnson and Ray (2004); Truccolo et al. (2005). However, the averaging

can be done in three different ways: averaging over observation time, averaging across trials and averaging across neurons. The choice of the averaging depends on the problem of interest and the limits of the experiment. For more details see Rieke et al. (1997), (Gerstner and Kistler, 2002, p.15-18) and the references therein. Notice that using the conditional intensity function (conditioning on the spiking history), Kass and Ventura (2001); Brown et al. (2003), also falls in the rate coding category.

2. *Temporal Coding* claims that the precise timing and coordination of spikes encode information in the brain, Theunissen and Miller (1995); Victor (1999); Panzeri et al. (2001); Reich et al. (2001); Dayan and Abbott (2001).

3. *Correlation/Synchrony Coding* states that the correlations between the spike trains summarize information, and that the information can be captured in the synchronous firing of a population of neurons Johnson (1980); Gray et al. (1989); de Charms and Merzenich (1996); Borisyuk and Borisyuk (1997); Singer (1999); Averbeck et al. (2006); Gouwens et al. (2010).

4. *Phase Coding* captures the information through the time of firing relative to a baseline periodic wave, O'Keefe and Recce (1993); Harris et al. (2002); Kayser et al. (2009).

Of course, there have been studies on combining different codes together to come up with a better understanding of the neural spike trains, Kass and Ventura (2001). In Chapter 4 we propose a new methodology in the analysis of spike trains which, in some ways, combines the rate code and a temporal code.

Although rate coding ignores any information possibly encoded in the temporal structure of the spike train (such as the time it takes to reach the first spike after the stimulus onset), rate coding is highly robust with respect to the interspike interval noise, Kostal et al. (2007); Stein et al. (2005). Following a point process approach, our main interest falls in the rate coding category, however, Chapter 4 will introduce a model which borrows properties from both the rate and temporal codes.

The literature on rate coding techniques is rich. In addition to the aforementioned studies, Brown et al. (2001) developed an adaptive filter algorithm to track neural receptive field plasticity (the change in the neural function and neural connections by learning). The work of Dimatteo et al. (2001), introduced a Bayesian approach (BARS) in spline curve-fitting with free knots. Kass and Ventura (2001) introduced the Inhomogeneous Markov Interval (IMI) process based on the conditional intensity

function of the underlying stochastic process governing spiking activity. According to Kass and Ventura (2001) "The IMI processes avoid the assumption that the spike trains are Poisson processes, which fails to account for effects such as refractory period." Kolaczyk and Nowak (2004) and Kolaczyk and Nowak (2005) developed a multiresolution approach for likelihoods based on which one can estimate the mean function of the spiking activity, Ramezan et al. (2014). Truccolo et al. (2005) derived a GLM-type model for the conditional intensity function of the underlying Poisson process governing the spiking activity. They studied the relationship between the neuronal activity and several covariances including the spiking history of the neuron. For more details on neural coding techniques refer to Stein et al. (2005); Gerstner and Kistler (2002); Rieke et al. (1997) and the references therein.

The first use of the term "information" to describe the content of the neural action potentials occurred in the book authored by Adrian (1928). Our purpose is to find possible methods through which the brain interprets the information content of the spike trains; however, to do this, one should have a good understanding of neural *information* before studying the information coding methods. According to (Dayan and Abbott, 2001, p.123), it is important to investigate "how much ... the neural response tell[s] us about the stimulus?" They use information theory tools, such as entropy, to explain the information content in the spiking activity. Given a stimulus s, suppose that $\Pr(r|s)$ shows the probability of a response at rate r given that the stimulus is fixed at s, and let $\Pr(r)$ show the marginal distribution of the response rate r. Dyan and aboot (Dayan and Abbott, 2001, p.127) define the *mutual information* as

$$I_m = \sum_{r,s} \Pr(s)\Pr(r|s) \log_2 \Big(\frac{\Pr(r|s)}{\Pr(r)} \Big). \tag{1.5.1}$$

Expanding this equation we get,

$$I_m \;\; = \;\; \sum_{r,s} \Pr(s)\Pr(r|s) \log_2 \Big(\Pr(r|s) \Big) - \sum_{r,s} \Pr(s)\Pr(r|s) \log_2 \Big(\Pr(r) \Big).$$

Therefore, the mutual information is, in fact, the difference between the entropy of the total response (under repeated presentation of the stimulus), and the total entropy under different stimuli for different trials. This difference is the entropy associated with that part of the response variability that is not due to the changes in the stimulus. For the details of derivation of Equation 1.5.1 refer to (Dayan and Abbott, 2001, p.126-128). For a detailed study on methods of measuring mutual information see Chechik (2003); Rieke et al. (1997). Gerstner and Kistler (2002) also used the same framework to measure the information content in neural activity. The work of Rieke and his colleagues (Rieke et al., 1997, Ch.3), provided details about why information theory is

a powerful tool in this context. For a review of the applications of the information theoretic tools in the neural activity refer to Victor (2006).

An interspike interval can be anywhere from a few to tens of milliseconds, increasing the spike rate of an individual neuron to more than a few hundred spikes per second. Neuroscience studies often involve simultaneous recording of many neurons on each of the several experimental subjects which are being monitored during different trials over periods of hours over days and days of experiments. This results in significantly large datasets, therefore, any estimation algorithms should be computationally efficient. Following a rate code approach, the interesting method of BARS, Dimatteo et al. (2001), utilizes the Reversible Jump MCMC method, Green (1995), to estimate the number of the knots, and to locate them efficiently, which makes it computationally intensive. According to Hastie and Tibshirani (1990), besides the computational cleanness when the knots are given, choosing the number and the location of the knots is a challenging problem. The multiscale model which is used in this dissertation, Kolaczyk and Nowak (2004, 2005), is computationally efficient and easy to implement.

1.6 Statistical analysis of neural spike trains

Statistical analysis of spike trains dates back to the work of Brink et al. (1946), where they provided a histogram-based approximation of the pdf of interspike intervals (ISIs) of a frog's neuron. According to Moore et al. (1966), due to the variability in the interspike intervals, the underlying process of the neural spiking activity is stochastic. The variability and the randomness in neural responses to repeated stimuli, Joeken et al. (1997), as well as the instantaneous nature of the action potentials, for which they are called spikes, motivate stochastic point processes as the underlying statistical framework for the analysis of neural spike trains. In particular, some of the common approaches in statistical analysis of neural spike trains are Integrate-and-Fire (IF) models, point processes and GLM-based models, filtering (smoothing), and renewal processes. Working with data, the likelihood method for estimation is a popular estimation technique for each of these approaches, Barbieri et al. (2001); Kass and Ventura (2001); Brown et al. (2003); Paninski (2004); Paninski et al. (2004, 2008).

1.6.1 Integrate-and-Fire model (IF)

The Integrate-and-Fire model for spike trains, Gerstein and Mandelbrot (1964); Stein (1965); Burkitt (2006a,b), is a dynamical model, which can take into consideration biological aspects such as refractoriness, all-or-none spiking law (see Section 1.3.1), as well as sub-threshold electrical activity (the electrical activity across the cell's membrane which is under the threshold of excitation). A simple case of an IF model is defined by the following stochastic linear dynamics, (Paninski et al., 2008, ch.10).

$$dV(t) = \Big(g(t)V(t) + I(t)\Big)\, dt + \sigma\, dB_t\,, \tag{1.6.2}$$

where $V(t)$ is the membrane potential, $I(t)$ is the input current to the cell, $g(t)$ is the conductance of the membrane and the stochastic term B_t is a standard Brownian motion. Every time $V(t)$ passes the threshold of excitation, $V_{threshold}$, a spike is initiated, after which $V(t)$ is reset to $V_{reset} < V_{threshold}$. This process makes the spike generation a threshold-crossing problem. Paninski and his colleagues, (Paninski et al., 2008, ch.10), discuss IF models from three different points of view; diffusion processes, state space (hidden Markov) models, and point processes. Burkitt (2006a), (Paninski et al., 2008, ch.10) discuss parameter estimation of the IF model using a likelihood-based approach. Equation (1.6.2) can be generalized in many ways, including having different forms for the conductance or the voltage function. To incorporate stimulus signal effect, Paninski et al. (2004) models $I(t)$ as a function of the stimulus signal $x(t)$.

IF models are popular due to their simplicity and computational efficacy. However, this computational power comes at a price. They do not fully model the dynamic behaviour of real neurons, (Brunel, 2010, ch.7). On the other end of the spectrum of conductance-based methods are Hodgkin-Huxley models, which are more complex and computationally intensive, but are closer to the real dynamic behaviour of neurons. The original Hodgkin-Huxley model, Hodgkin and Huxley (1952), has been significantly generalized since 1952 to more complex models with more components, Borg-Graham (1999); Purvis and Butera (2005), incorporating more ion channels than the initial sodium and potassium channels.

The conductance-based dynamical system models, i.e. IF or Hodgkin-Huxley models, are at a finer time scale than the spike trains. However, it has been shown that certain types of point process models fit well to the data generated from an IF model, Paninski et al. (2008); Komoya and Kass (2008); Paninski et al. (2007). This shows that the point process approximations can indirectly provide insight about the underlying biophysiology, Paninski et al. (2008). More recent studies have made direct connections between the two approaches. Meng et al. (2011) introduces a sequential Monte

Carlo algorithm that combines the future and past spiking information to estimate the parameters of the dynamical system of a Hodgkin-Huxley model.

1.6.2 Point processes, histogram-based, and filtering models

Let $N(t)$ show the number of spikes up to, and including time t. The conditional intensity, or the history-dependent rate function of neural spiking activity $\theta(t|H_t)$ is defined to be

$$\theta(t|H_t) = \lim_{h\to 0} \frac{\Pr\Big(N(t+h) - N(t) = 1 \,\Big|\, H_t\Big)}{h}, \tag{1.6.3}$$

where H_t is the history of the spike train up to time t, i.e. the spike times up to time t. In survival analysis, this function is known as hazard function, Kalbfleisch and Prentice (1980). Based on a given spike train $t_1, \cdots, t_n$ on the observation interval $(0, T)$, the probability density of exactly these n spikes in $(0, T)$ in a point process framework is, Brillinger (1988); Brown et al. (2003),

$$\begin{aligned} \Pr(t_1, ..., t_n) &= \prod_{i=1}^{n} \theta(t_i|H_{t_i}) \, \exp\Big\{ - \int_0^T \theta(t|H_t)\, dt \Big\} \\ &= \exp\left\{ \int_0^T \log \theta(t|H_t)\, dN(t) - \int_0^T \theta(t|H_t)\, dt \right\}. \end{aligned} \tag{1.6.4}$$

We will use this likelihood function later in Chapter 2. Many research studies suggest that the conditional intensity function can summarize information content of neural spike trains, Brown et al. (2001); Barbieri et al. (2001); Truccolo et al. (2005); Chen et al. (2009); Sarma et al. (2010); Banerjee et al. (2012), and there are different methods in the literature for estimating the history-dependent spiking rate function. Barbieri et al. (2001) studied this problem under inhomogeneous Gamma, Inverse Gaussian and inhomogeneous Poisson models. Under the inhomogeneous Poisson process framework, neural spiking activity is independent of the history, thus H_t can be omitted from Equation (1.6.3), i.e. $\theta(t|H_t) = \theta(t)$. Although some of the assumptions of Poisson process are not justified due to biological phenomena such as refractory period, burst spiking activity or temporal dependence among spikes, Poisson models are still widely used in the literature, Hanes et al. (1995); Banerjee et al. (2012). Some studies suggest adjustments for refractory period, (Brunel, 2010, p. 163). More recently, there have been studies in which, particular areas of the brain have been claimed to show Poisson variability in their spike counts, Softky and Koch (1993); Maimon and Assad (2009); Averbeck (2009). However, caution should be used when employing these results. We will show in Chapter 3 that spike count variability can depend on the time scale and

the bin size.

Incorporating explanatory variates, Truccolo et al. (2005) discretized the observation window into very small subintervals and estimated the conditional intensity using GLMs for Bernoulli trials. GLM models for the conditional intensity function have been widely used in the literature, from retinogeniculate cell (RGC) responses, Ahmadian et al. (2009), to auditory cortex models, Calabrese et al. (2011), primary motor cortex, Truccolo et al. (2008) and spatio-temporal correlation analysis in the visual system, Pillow et al. (2008). All of these applications include, in one way or another, formulating the spiking activity $\theta(t|H_t)$ as a function of the stimulus signal and/or some explanatory variables (e.g. spiking history), Stapleton et al. (2006); Brown et al. (1998); Carandini et al. (2007); Calabrese et al. (2011).

Incorporating history in Equation (1.6.3), Kass and Ventura (2001) introduced the Inhomogeneous Markov Interval (IMI) process, which formulates the conditional intensity function as a function of the experimental clock, and the time elapsed from the previous spike. They assume that

$$\theta\big(t \,|\, H_t\big) = \theta\big(t \,,\, t - s_*(t)\big) \tag{1.6.5}$$

where $s_*(t)$ is the time of the last spike preceding t. In particular, they use the product form of Equation (1.6.5) where

$$\theta\big(t \,|\, H_t\big) = \theta_1(t)\, \theta_2\big(t - s_*(t)\big). \tag{1.6.6}$$

Through data analyses, they show that this model outperforms the homogeneous Poisson process, proving that spiking activity of neurons is not "memoryless," at least in the supplementary eye field. The IMI model combines the effect of the experimental clock and the spiking history in a multiplicative way. Paninski et al. (2008) discusses two other options. A Time Rescaling Renewal Process (TRRP) transformation of $t - s_*(t)$, i.e. substituting it by $\Theta_1(t) - \Theta_1(s_*(t))$ where $\Theta_1(t) = \int_0^t \theta_1(u)\, du$. The other method is to combine t and the history in an additive way, where they show that the additive IMI model is, essentially, equivalent to a soft thresholding IF model.

It is important to note that the Poisson process is a good approximation for the analysis of neural spike trains when the data is pooled across many independent trials, Daley and Vere-Jones (2008). In this case, histogram techniques, such as the Peri-Stimulus Time Histogram (PSTH), are a good visualization technique. The PSTH is, essentially, a histogram of pooled data across trials with a repeated stimulus. Figure 1.7 shows the PSTH and the raster plot (common visualization methods for raw data) of 129 trials

of retinal cell data in a vision experiment on connected retinal ganglion cell (RGC) and lateral geniculate nucleus (LGN) neurons, Carandini et al. (2007); Sincich et al. (2007). The raster plot is the plot of actual spike times in an experiment. Each row in this plot shows the spike times of one trial. The details of the experimental conditions for this dataset are discussed in Chapter 3.

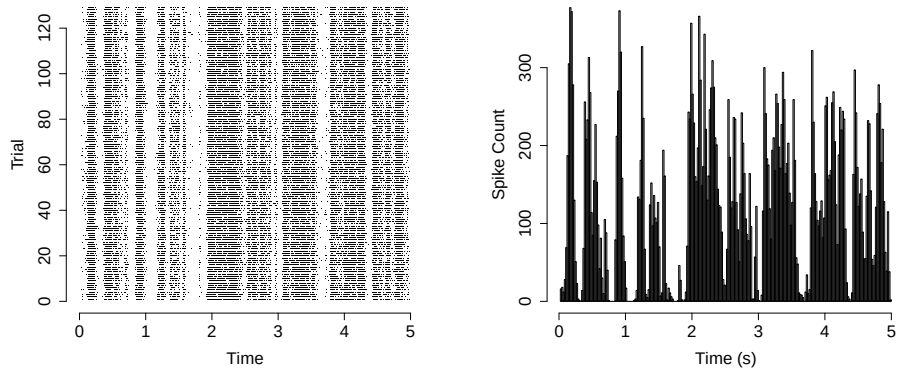


Figure 1.7: *The raster plot and the PSTH based on 129 trials of repeated stimulus on an RGC neuron.*

The Bayesian Adaptive Regression Splines (BARS) method of Dimatteo et al. (2001) can be used to smooth the PSTH and to estimate the intensity function of spiking activity. Although its performance in identifying the change points works relatively well, BARS is computationally expensive.

1.6.3 Filtering/Smoothing models

Both the BARS and GLM models discussed here lie within the more general framework of *filtering* or *smoothing* in the signal processing literature. Although no optimal filter design exists to perfectly decode spike train data, it is still possible to design good firing rate filters with some optimal characteristics, Paulin (1992). As a result, filtering techniques are widely used in the statistical analysis of neural spike trains including, but not limited to, the control of prosthetic devices, Brockwell et al. (2004); Srinivasan et al. (2007), estimating the location of a freely moving rat from the activity of hippocampal place cells, Brown et al. (1998), and detecting patterns in responses to visual stimuli, Carandini et al. (2007).

In a Poisson process framework, Brown et al. (1998) introduced a modified Bayes filter to estimate the location of a rat based on the firings of hippocampal place cells, which is, essentially, the MAP estimate derivation of the Kalman filter. Using kernel estimation methods in the framework of an inhomogeneous Poisson process, Lehky (2010) suggests minimizing a bias squared error criterion to find optimal width for the Gaussian kernel. Kass et al. (2003) compares Gaussian filters, cubic splines and BARS, where they argue that BARS outperforms the other two methods when the neural firing rate changes rapidly over time.

The Kalman filter has been extensively used in neural spike train data analysis. In a study involving hippocampal place cells recordings from a freely moving rat, Barbieri et al. (2004) modelled the rate function of an inhomogeneous Poisson process as a function of the coordinates of the animal's position and the coordinates of the place field's centre. They developed a Kalman filter to predict the location of the animal on the disc. Discretizing time into small time bins and modelling spike trains as realizations of Bernoulli trials, Eden et al. (2004) develops point process filter analogues of the Kalman filter. It is noteworthy that since the Kalman filter only uses the present state of the process and the result of the calculations in its past steps, it can be used in real-time in the analysis of neural data. This makes it attractive for neuroscience research, particularly in neuro-prosthetic applications, Wu et al. (2006); Srinivasan et al. (2007). For more applications refer to Srinivasan et al. (2007); Paninski (2010).

1.6.4 Renewal processes

The Interspike Interval (ISI) distribution has usually a much heavier tail than exponential distribution, Brown et al. (2002, 2003); Truccolo et al. (2005). This motivates utilizing other ISI distributions, and hence, other renewal processes than the Poisson process, Perkel et al. (1967); Stein (1967); Johnson (1996); Nawrot et al. (2008); Grün (2009). Beside the inhomogeneous Poisson process, the most common renewal processes in the analysis of spike trains are Gamma, Log-Normal and Inverse Gaussian. In Chapter 4 we introduce Skellam Process with Resetting (SPR). There, interspike intervals are modelled as realizations of inter-record intervals of the difference between two independent Poisson processes with an additional adjustment for refractory period. Figure 1.8, adapted from van Vreeswijk (2010), shows the ISI distribution based on Gamma, Log-Normal and Inverse Gaussian distributions. The spike rate is held fixed at 50Hz and the coefficient of variations are 0.5 (black), 0.75 (red), 1 (green) and 1.5 (blue).

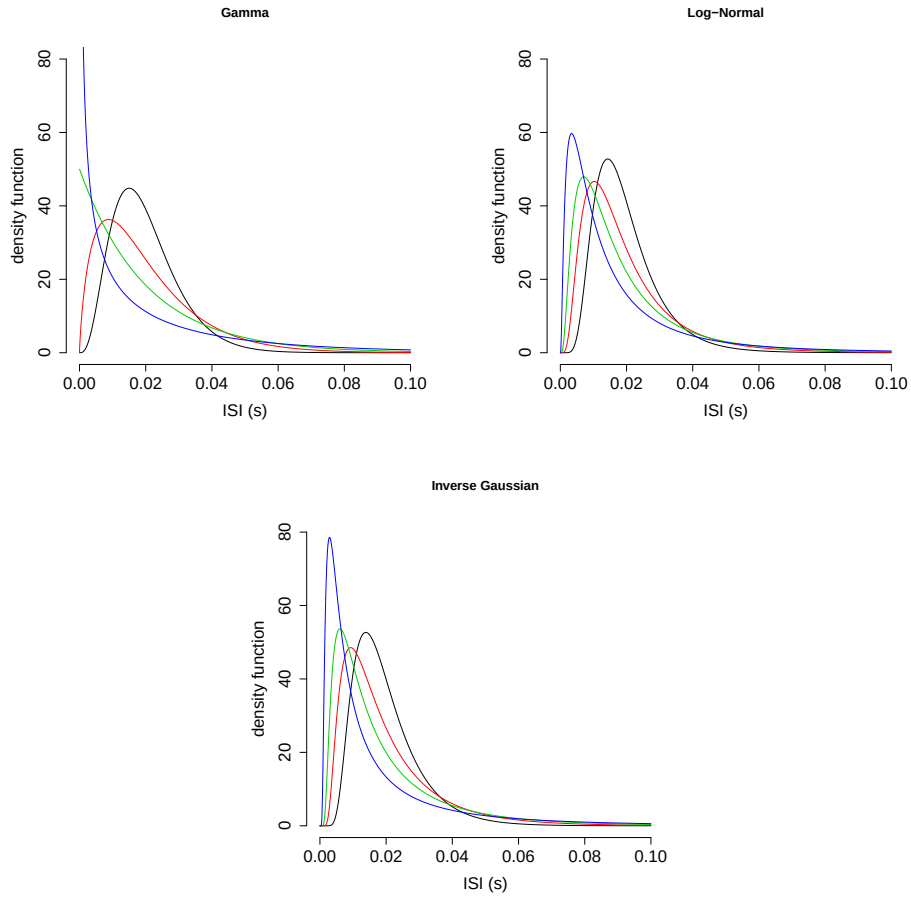


Figure 1.8: *ISI distributions for Gamma, Log-Normal and Inverse Gaussian renewal processes. The spiking rate is held fixed at* $R = 50Hz$ *and the coefficient of variation (CV) is 0.5 (black), 0.75 (red), 1 (green) and 1.5 (blue).*

All these cases assume stationarity, which implies that the parameters of the ISI distribution are constant over time, which is not a feasible assumption in real data analysis. To address this problem, it is necessary to introduce time-dependent versions of the aforementioned renewal processes. Barbieri et al. (2001) introduces inhomogeneous Gamma (IG) and inhomogeneous Inverse Gaussian (IIG) probability models for the interspike distribution. Let $\{t_1, t_2, \ldots, t_n\}$ denote the spike times and let $\theta(t)$ be a strictly positive intensity function. The IG and IIG models are respectively defined as,

Barbieri et al. (2001),

$$f_t(t_k|t_{k-1}) = \frac{\gamma\,\theta(t_k)}{\Gamma(\gamma)}\left(\gamma\int_{t_{k-1}}^{t_k}\theta(u)\,du\right)^{\gamma-1}\exp\left\{-\gamma\int_{t_{k-1}}^{t_k}\theta(u)\,du\right\} \quad (1.6.7)$$

$$f_t(t_k|t_{k-1}) = \frac{\theta(t_k)}{\sqrt{2\pi\left(\int_{t_{k-1}}^{t_k}\theta(u)\,du\right)^3}}\exp\left\{-\frac{1}{2}\frac{\left(\int_{t_{k-1}}^{t_k}\theta(u)\,du-\psi\right)^2}{\psi^2\int_{t_{k-1}}^{t_k}\theta(u)\,du}\right\} \quad (1.6.8)$$

where Γ is the gamma function, γ and ψ are, respectively, the constant parameters of IG and IIG probability models, and $\theta(t)$ is the time varying parameter. Notice that if $\gamma = 1$, IG is the inhomogeneous Poisson probability model. Modelling CA1 hippocampal pyramidal neurons, Barbieri et al. (2001); Brown et al. (2002) showed that both IG and IIG models outperform inhomogeneous Poisson process in terms of fit to the data.

Although the distribution of real interspike intervals has a much heavier tail than the exponential distribution, Gabbiani and Koch (1998) showed that if the threshold potential after each spike in the integrate-and-fire model is reset to a random value according to an exponential distribution, the corresponding spike train will follow a Poisson process. The work of Farkhooi et al. (2009) reports negative autocorrelation (a.k.a. serial correlation) in some parts of mammalian brains and fish, which violates the assumptions of the renewal process. They suggested an alternative autoregressive point process model to incorporate this autocorrelation among interspike intervals. For more history-dependent models (e.g. Markov Point Processes) refer to Johnson (1996); Kass and Ventura (2001); Truccolo et al. (2005); Paninski et al. (2007).

1.6.5 Likelihood-based inference

The neural encoding is a map between the stimulus space and the spike train space. Let $(0, T)$ denote the observation interval. Based on the spike train $\mathbf{t} = \{0 < t_1, ..., t_n < T\}$ and the stimulus x, the probabilistic formulation of the neural code is $\Pr(\mathbf{t}|x)$. Maximum likelihood estimates have desirable (asymptotic) properties, (Lehmann, 1983, ch.6), and the aptness of likelihood-based models in the analysis of spike trains has been addressed in the literature, Brown et al. (2003); Kass et al. (2005). These reasons motivate the extensive use of likelihoods in the neuroscience literature.

The work of Brillinger (1988) is one of the early and influential studies on utilizing likelihood functions in the analysis of neural spike trains. Incorporating the neural

integration phenomenon, he formulated the interaction between two and three neurons via likelihood functions in a point process framework. Brown et al. (2003) studies the ISI of hippocampal place cells under Gamma and Inverse Gaussian models and discusses some model selection criteria. Pillow (2007) studied the details of ML estimation in the point process framework for IF and GLM models. He presented log concave likelihood models and argued that optimization in a convex parameter space is efficient. Pillow et al. (2010) introduces log concave likelihood models for decoding information content of spike trains about the stimulus signal. For more details and examples on the use of likelihoods in the context of neural spike trains refer to Paninski (2004); Paninski et al. (2004); Salimpour et al. (2011).

1.6.6 Nonparametric and Bayesian inference

In addition to the likelihood-based inference, nonparametric and Bayesian models have also been referred to in the neuroscience literature. Nonparametric methods do not impose distributional assumptions on neural data. Bootstrap and nonparametric regression are two common nonparametric statistical methods widely used in neuroscience, Kass et al. (2005). Nonparametric bootstrap confidence intervals will be used throughout this dissertation. Kaufman et al. (2005) suggested a spline-based generalized non-parametric regression model to estimate the neural firing rate. Bootstrap and nonparametric regression are not the only nonparametric techniques employed in neuroscience contexts. Gourévitch and Eggermont (2007) proposed a nonparametric burst detection algorithm based on the ranks of the interspike intervals. Using Fourier and wavelet transforms, Nedungadi et al. (2009) introduced a nonparametric Granger causality to study the network connectivity in the brain from multivariate neural data. The nonparametric models have also been used to estimate neural firing rates.

Bayesian methods are also among the popular approaches in the analysis of neural spike trains, particularly in the decoding problem where interest lies in estimating the probability function of a stimulus x given the spike trains $0 < t_1, ..., t_n < T$. We have,

$$\Pr(x|\boldsymbol{t}) \propto \Pr(\boldsymbol{t}|x) \times \Pr(x)\,. \tag{1.6.9}$$

The first term in the right hand side of Equation (1.6.9) is the likelihood function of the spike trains, and the second term is the prior distribution of the stimulus signal. Koyama et al. (2010) develops a Bayesian state-space model to solve the decoding problem in a motor cortex (hand motion) study. In another study on the primary motor cortex in awake behaving monkeys, Wu et al. (2006) used a Bayesian Kalman filter model to compute the posterior probability of the hand motion conditioned on

a sequence of observed firing rates. In a Poisson process framework, Lehky (2004) developed Bayesian models for the estimation of mean and the variance of spike counts where they showed that relative to the simple likelihood analysis, their Bayesian models provide more precise interval estimates of responses.

1.7 Neural data collection

Neural data collection is an integral part of neuroscience research. The type of data to be collected in a specific research study depends on the objectives of the research. This section briefly discusses some of the common data collection methods.

Whether to collect the data from an ensemble of nerve cells including thousands or millions of neurons, or to record single neurons, an approach which considers the contribution of the individual cells to the network of the nerve cells, is a controversial matter in neuroscience research. While monitoring large populations of neurons measures the integrated neural activity of the population, single cell recordings provide details about the contribution of each individual cell to the behaviour, Criado et al. (2008). Furthermore, single cell recording methods are significantly more expensive and harder to implement as they require brain surgeries or other invasive methods. According to Windhorst and Johansson (1999), to examine the quality of the neural data, one should consider the following specifications:

1. "Detection sensitivity:" High sensitivity of the recording method is required to make useful measurements,

2. "Signal-to-noise ratio:" The ratio of the signal power to the noise power, which should be large enough to make experimentally useful conclusions,

3. "Spatiotemporal resolution:" The *spatial resolution* is the ability to discriminate between two points in space and *temporal resolution* is the ability to discriminate between two points in time, Slavin and Bluemake (2005). These values should be large enough to answer the experimental questions posed.

4. "Fidelity:" Accurate reflection of the underlying physiological phenomena by the data and no disruption by the recording methodology itself.

For details about data collection techniques and the related issues refer to Windhorst and Johansson (1999).

1.7.1 Recordings from populations of neurons

According to (Carlson, 2007, p.148), *Computerized Tomography (CT)*, which is generally referred to as the *CT scan*, was the first method developed to monitor the anatomy of a living brain. The CT scan machine emits x-rays to the head, measuring the amount of radioactivity passing through it. These measurements are performed from many different angles, allowing the machine to produce pictures of the skull and its contents.

A more complicated and accurate method, developed after the CT scan, is *Magnetic Resonance Imaging (MRI)* which works based on properties of the cells being exposed to a strong magnetic field. Unlike the CT scan, MRI is not an x-ray based method, which makes it less invasive, (Carlson, 2007, p.149). Additionally, MRI generates pictures of the lateral views and the front view of the brain, which are not possible with CT scans.

The most recently introduced brain imaging method is the *functional MRI (fMRI)* which, unlike the MRI, can also monitor the function of the brain, (Carlson, 2007, p.155). It is a blood-oxygen-level dependent method (BOLD). The BOLD signal measures the changes in the ratio of oxygenated to deoxygenated blood, which is an index for brain activity, (Kandel et al., 2000, p.374-375).

Magnetoencephalography (MEG) is another magnetic field based neuroimaging method. It allows for inferences about the location of the dipole that gives rise to a specific magnetic field, (Zigmond et al., 1999, p.152), which is impossible with the other methods. MEG is usually used in cognitive research studies. It measures extremely weak fields outside the head, and can pick up fields associated with concerted actions of a few thousands of neurons, Windhorst and Johansson (1999).

One of the most commonly used methods of neural data collection is *electroencephalography (EEG)* which is a method of recording the electrical activity of the brain through scalp electrodes. It measures the integrated potentials of a population of neurons. According to (Windhorst and Johansson, 1999, pp.971-995), beside the relatively lower costs of the equipments, one of the major advantages of EEG is the ability to measure electrical changes in the brain, which is not possible with the fMRI method. Another advantage of EEG is its ability to record the periodic neural activities of the brain discussed in Section 1.3.1.

1.7.2 Single cell recordings

Optical methods and recordings from surgically implanted electrodes (microelectrodes) are two of the data collection methods at the single cell resolution level, Windhorst and Johansson (1999). Both methods include intracellular and extracellular recordings, but this dissertation is based on the extracellular measurements recorded by surgically implanted electrodes. The earliest formal publications on single cell recording using surgically implanted electrodes are Adrian (1928, 1930) where the electric discharge of sensory neurons in frogs, cats and rabbits are measured.

Optical methods involve using a *dye* which is calcium-sensitive or voltage-sensitive. One typical optical method is that after the indicator is injected, photons are emitted to the area of interest in the brain, whose spectral properties reflect the neural activity. For details on these methods refer to (Windhorst and Johansson, 1999, Ch. 4, 16, 34) and Antic et al. (1999). For developments on calcium imaging methods see Takahashi et al. (2007).

Intracellular recordings measure the cross membrane potential of neurons by using the *patch clamp technique*, which involves sealing a glass or quartz pipette onto the membrane. These methods are used to study fine-level behaviour of neurons, e.g. ion channels dynamics. For details on techniques and challenges of the intracellular recordings refer to Brown and Flaming (1986); Sherman-Gold (1993).

Extracellular recordings, which are used to extract data employed in this dissertation, measure the electrical field potential outside neurons. The methods for such recordings are used mainly for detecting action potentials or spikes. Extracellular recordings are widely used for behavioural studies, particularly in freely moving animals, O'Keefe and Dostrovsky (1971); Carandini et al. (2007); Maimon and Assad (2009); van der Meer and Redish (2009); Walker et al. (2011). With the development of multielectrodes, Taketani and Baudry (2006), these recordings allow for collecting electrical activities from a large ensemble of neurons. However, compared to intracellular recordings, extracellular recordings are harder to interpret because they are not direct measurements of the membrane potential. The first extracellular recording results on freely moving animals was published in Strumwasser (1958), where ground squirrels were studied while running around or sleeping. For the details on different techniques and types of electrodes used in extracellular recordings, refer to McNaughton et al. (1983); Gray et al. (1995); Wise (2007). The PhD thesis of Ferguson (2011), discusses different methods, challenges, sources of error and current state-of-art of extracellular as well

as intracellular recordings. There, as well as in the references therein, the relationship between the intracellular and extracellular recordings has also been addressed.

The spike trains we use in this dissertation are all collected through extracellular recordings using surgically implanted microelectrodes. These thin electrodes have fine enough tips to indirectly measure the electrical activity of individual neurons. The procedure of implanting the tip of an electrode into a neural area of the brain is called *stereotactic surgery*, (Carlson, 2007, pp.138-141). A *stereotaxic atlas*, which is a collection of drawings of sections of the brain of a particular animal along with their corresponding landmarks on the skull, provides precise coordinates for stereotactic surgery, (Carlson, 2007, pp.139). Neuroscientists use the skull landmarks in the stereotaxic atlas to implant the microelectrodes into the corresponding areas of interest in the experimental subject's brain. Figure 1.9 provided by Dr. Matthijs van der Meer of the Biology Department, University of Waterloo, shows the recording unit containing a collection of the microelectrodes.

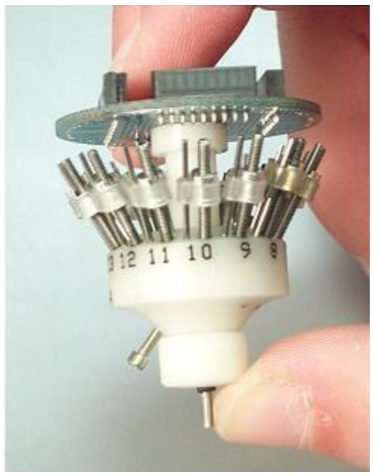


Figure 1.9: *The recording unit housing microelectrodes used in data collection.*

Attaching the unit to the brain, the electrical activities of the corresponding neurons are recorded using computers. After filtering noise, the number of neurons being monitored by the electrodes and their corresponding spike trains are determined through a procedure called *spike sorting*, Lewicki (1998); Pouzat (2008). According to (Windhorst and Johansson, 1999, Ch.41 & 42), the very high temporal and spatial resolution of microelectrode recording has made it the principle method for function and behavioral analyses of neurons. After the recording task is complete, the brain may be perfused and extracted to investigate the exact placement of the microelectrodes and

their recording sites.

As mentioned above, methods which have been discussed in Sections 1.7.1 and 1.7.2 have different spatiotemporal resolutions. The quality of the recording device as well as the design of the experiment affect the spatiotemporal resolution. For example, the spatial resolution of the fMRI method depends highly on the strength of the magnet being used in the machine. The stronger the magnet, the better the spatial resolution. Figure 1.10 provided by Dr.van der Meer, shows the spatiotemporal resolutions of the data collection methods discussed in this section.

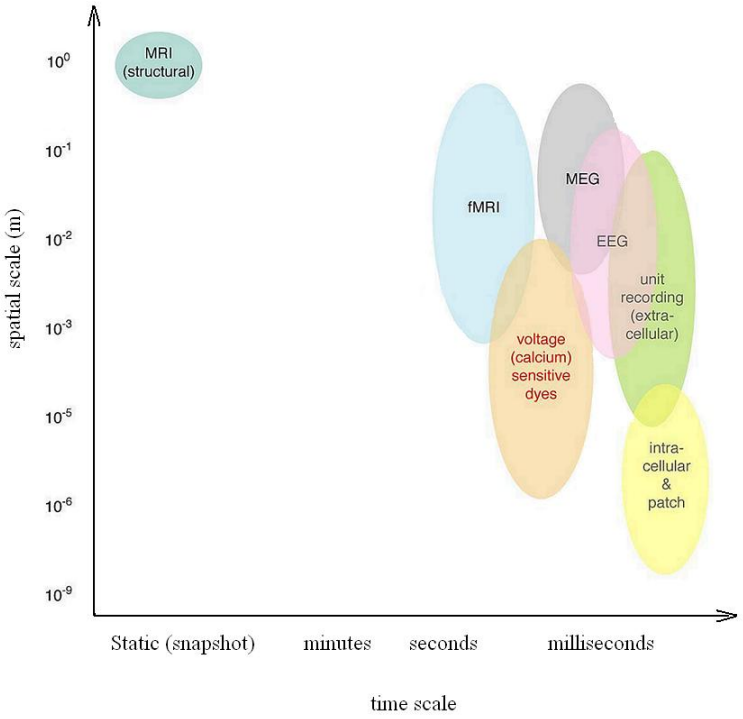


Figure 1.10: *The spatiotemporal resolutions of neural activity recording methods.*

The methods noted here are only some of the techniques used to record neural activity. For more details about data collection methods and techniques, refer to Windhorst and Johansson (1999); Martin (1998).

1.8 Discussion

This chapter provided the background knowledge in the neuroanatomy and neurophysiology of the nervous system, which is needed for the rest of the thesis. Neural

spike trains were introduced and the common statistical methods (IF, point processes, histogram-based, filtering, renewal processes, likelihood-based, Bayesian and nonparametric inferences) for the analysis of such data were discussed. The data collection techniques, both single cell recordings and recording the activity of an ensemble of neurons, were also discussed. In the remainder of the thesis, different statistical models for the analysis and modelling of the neural spike trains will be developed.

1.9 Online data resources

Below is a list of online sources for neural spike trains:

- http://www.tech.plymouth.ac.uk/infovis/LAB_Downloads.htm
- http://gaya.jp/data/
- http://bmi.neuroinf.jp/
- http://www.carmen.org.uk/portal
- http://www.hirnforschung.net/cneuro/
- http://crcns.org/
- http://neurodatabase.org
- https://sites.google.com/site/spiketrainanalysiswithr/Home/analysis-gallery
- http://www.biomedicale.univ-paris5.fr/physcerv/C_Pouzat/Data.html
- http://neuinfo.org/

It should be noted that the second website provides data collected based on the functional multineuron calcium imaging (fMCI) technique, which is an optical data collection method with a spatial resolution at the single cell level, Takahashi et al. (2007).

Chapter 2

Multiscale Analysis of Neural Spike Trains

2.1 Introduction

This chapter studies the multiscale analysis of neural spike trains, through both graphical and Poisson process approaches. The powerful interspike interval (ISI) plot, Ramezan et al. (2014), which simultaneously visualizes characteristics of neural spiking activity on different time scales, is employed. We propose multiscale estimates of the intensity functions of spike trains with additive or multiplicative periodic components to address the periodicity of the spike trains caused by a stimulus signal or by brain rhythms. A cross-validation scheme is provided to choose the tuning parameters of the multiscale model, and its unbiasedness and robustness is studied.

2.2 Statement of the problem

As discussed in Chapter 1, neurons communicate with each other through spike trains. How spike trains code information in the brain is of great interest, but its complete characterization is still far from being settled, (Gerstner and Kistler, 2002, p.14), Brown et al. (2004); Grun and Rotter (2010). *Rate coding* (see Chapter 1) is one of the most common ways of neural information coding considered in the literature. It states that the information in a spike train can be captured by counting the number of spikes in a short period of time, Brown et al. (1998, 2002); Truccolo et al. (2005). In short, the stronger the stimulus, the higher the firing rate and vice versa, Gerstner and Kistler (2002). Therefore, proper estimation of the rate of spiking activity can provide valuable information about the stimulus signal, to which the neuron is exposed.

When studying spike trains as realizations of point processes, one popular approach is to model spike trains as realizations of an inhomogeneous Poisson process $\{N(t), t \geq 0\}$ with a time varying intensity function $\theta(t)$. Here, it is implicitly assumed that it is the function $\theta(t)$ that codes information, Brown et al. (1998); Truccolo et al. (2005). The time varying intensity function $\theta(t)$ allows for modelling long term, temporary and/or sudden increase/decrease in spiking activity, which are usually seen in spike train data. These changes in the neural spiking rate can be caused by different neural phenomena from different time scales.

Neural phenomena happen at a variety of time scales. At the finest scale, the *refractory period* is in the order of a few milliseconds. Recall that the refractory period is a short time interval after each spike during which a neuron cannot fire a second spike. *Burst spiking* activity, which is the sudden and frequent spiking activity of neurons with short interspike intervals, can happen on the scale of less than a second, Izhikevich (2000); Li et al. (2009); Tokdar et al. (2010). It is important to emphasize that the term "bursting activity" has also been associated with the activation of T-type calcium channels, Llinás and Steriade (2006). However, in this thesis, bursting and burst spiking activity only refer to short interspike interval and these terms will be used interchangeably. In contrast, long term potentiation and long term depression, which are compelling physiological models of learning and memory, are both examples of neural phenomena which should be studied in time scales of at least an hour, Beck et al. (2000); Cooke and Bliss (2006); Massey and Bashir (2007). These characteristics from multiple time scales show that neural activity is fundamentally a multiscale process. The work of Nelson, Nelson (2002), shows that the amount of variability in the spike trains depends on the time scale. The different time scales of these biological phenomena suggest the use of a multiscale approach for the analysis of neural data.

Although there have been discussions in the literature about the effect of time scale on spike train data analysis (e.g. Nelson (2002) from the interspike interval point of view or Kass and Ventura (2006)), to the best of our knowledge, there has not been any study tailored to accommodate phenomena from multiple time scales within a Poisson process framework. In particular, discretizing time to small bins, a common approach is to choose a *fixed* bin size (and according to Omi and Shinomoto (2011) usually in a subjective manner), and perform the analysis on the binned spike counts. Formal methods to choose a fixed bin size with respect to some "goodness-of-fit" criteria are discussed in Omi and Shinomoto (2011); Shimazaki and Shinomoto (2006). The bin-size may also be chosen based on a combination of biological facts and "goodness-of-fit"

criteria, Stapleton et al. (2006). In the current chapter, we show that making these bins of the same size is not appropriate, or efficient. We also provide a multiscale method for choosing "optimum" bin sizes across the observation window. We have noticed that the correlation between the spike counts of paired Retinal Ganglion Cell (RGC) and Lateral Geniculate Nucleus (LGN) neurons in a rhesus monkey study depends on the time scale, which reinforces the importance of a rigorous and proper choice of bin size. Using piecewise constant functions to estimate the intensity function of neural spike trains in an inhomogeneous Poisson process framework, we estimate the intensity function of the spiking activity in response to a stimulus signal. This problem is of great interest in neuroscience, as the estimate of the intensity function is a foundation of the rate coding approach. It can be used to perform prediction, Brown et al. (1998), or to investigate the relationship between the spiking activity and the stimulus signal to which the neuron has been exposed [see Figure 3.1 (a)]. In the latter case, it will be shown that the spike latency (time difference between stimulus onset and the first spike) can play an important role in estimating the intensity function. We propose models for the time varying intensity function $\theta(t)$, which consist of two parts: $c(t)$, which captures the multiscale structure of the spike train, and a second term, which consists of trigonometric functions to model periodicity. These two components are then combined in a multiplicative or additive fashion to form $\theta(t)$, the intensity function of the neural spike trains. The multiscale estimate is computationally efficient, Kolaczyk and Nowak (2004, 2005), and if tuned correctly, it captures structures and variability in the stimulus signal.

2.3 Data visualization: interspike interval (ISI) plots

Data visualization is one of the most powerful tools in statistical data analysis, however, it is challenging to visualize different properties of large datasets through informative graphs, Bajaj (1999); Unwin et al. (2006); Gorban et al. (2007); Martin and Urbanek (2008). A single neuron can fire a large number of times per second, for example, during bursting activity. On the other hand the complete data set may cover a period of several hours or days. Similarly, the subjects can have periods of several hours when asleep, but they can also respond to stimuli in only a few milliseconds. Due to this multiscale nature of the spiking activity, graphical displays need to enable visualization of the properties of biological phenomena from different time scales. Moreover, since inhomogeneous Poisson process is a common framework for the analysis of spike train data, it is useful to visualize the data in a way that ensures that departure from time homogeneity is reflected in the graphs. We employ the Interspike Interval (ISI) plot, Ramezan et al. (2014), which manages to simultaneously show large and small

time scales, whilst also being able to visualize the departure from time homogeneity in the spiking activity. In this plot, the x axis represents the time of the i^{th} firing, T_i, while the y axis shows the time between the i^{th} and the $(i-1)^{\text{th}}$ firing in the log scale, $\log(T_i - T_{i-1})$. Panels (a) and (b) of Figure 2.1 show ISI plots of the data from connected lateral geniculate nucleus (LGN) and retinal ganglion cells (RGC) of an anesthetized rhesus monkey during a vision experiment. Panel (c) plots data from a hippocampal place cell of a freely moving rat and panel (d) shows data from a different study on hippocampal neurons of a freely moving rat. The marks on the horizontal axes, also known as a *rug plot* (Vanables and Ripley, 1999, pp.134-135), represent actual spike time records.

Different structures in different time scales are visualized in Figure 2.1. Vertical streaks in the graphs are associated with burst spiking activities. The red vertical boxes show some of these vertical structures. In panel (c) many of these vertical streaks are visible inside the red box. This panel represents data from a hippocampal place cell and the short interspike intervals occur as the rat passes through the place field of the particular place cell whose data is visualized. The vertical empty streaks (white in this graph) show temporal neural inactivity, which are associated with the gaps in the rug plots. Some of these are shown in blue boxes in panels (a) and (b) of Figure 2.1. Reading panel (c), the pattern in the ISI plot is different under and above the bright green horizontal line. This corresponds to the fact that the burst spiking activity of the place cell happens at a time scale finer than roughly $\exp\{-0.6\} = 0.54$ seconds. In other words if the rat is not on the place field of the cell, the typical interspike interval is more than 0.54 seconds. The horizontal dark band in panel (d) between -2 and -2.5 in the vertical axis (inside the brown box) translates to $7 - 12$Hz in frequency domain, and shows some periodic spiking activity at this frequency range. This characteristic is known as θ-rhythm and is common among hippocampal cells. It seems that there exists a change-point at about the vertical dark green line in panel (d). It turns out that this point is where the rat leaves the track and is no longer running on the maze.

Notice that no observation in any of these plots has a y-axis value of less than -7. This is consistent with the refractory period, which is usually at least 1 millisecond. Reading panels (c) and (d), it is clear that due to the high density of the spikes during the observation interval, the rug plot is not capable of reflecting any information about the spike train.

It can be seen that the ISI plots in panels (a) and (b) are very similar, but different from panels (c) and (d) which are the data from different experiments on different areas

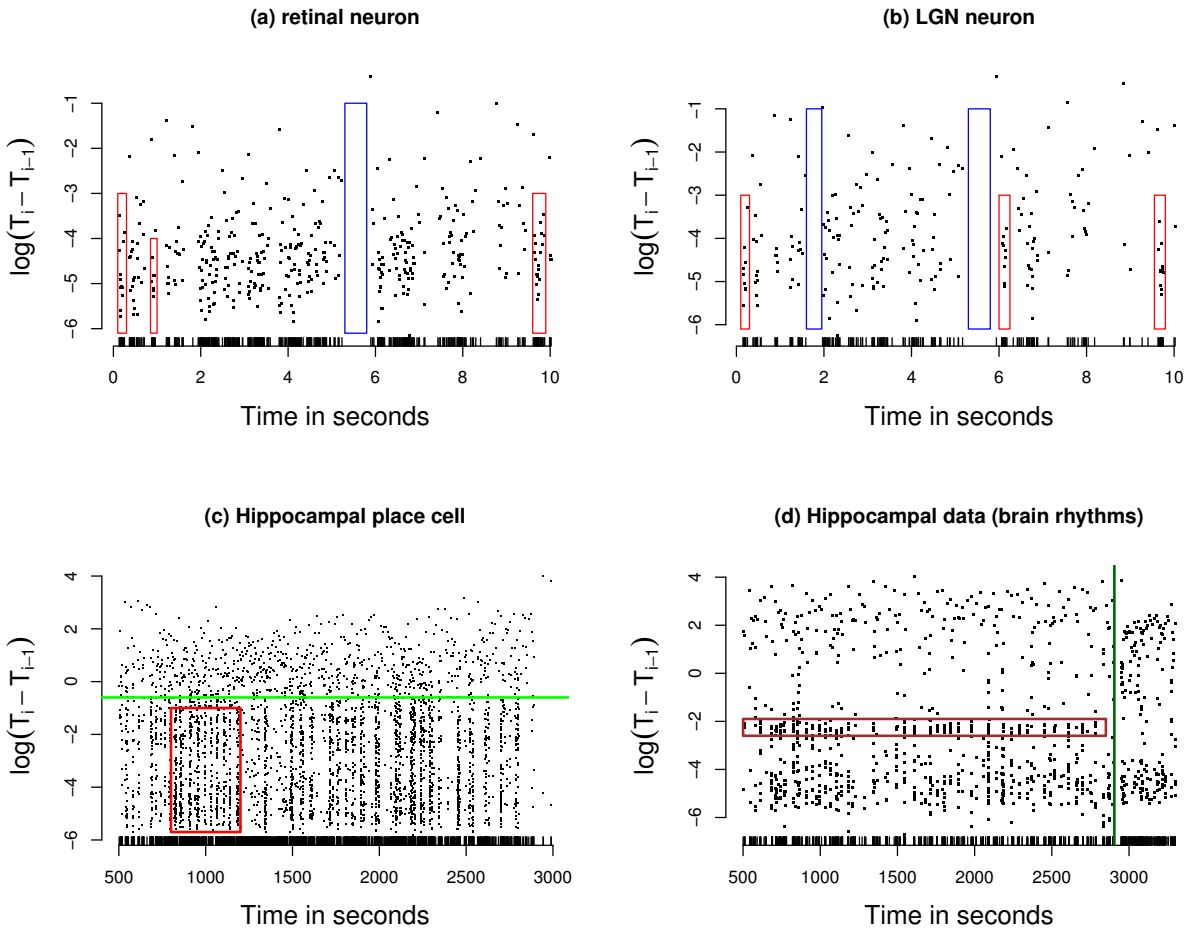


Figure 2.1: *Panels (a) and (b) show sample ISI plots of the spike train data collected from a retinogeniculate synapse i.e., neurons in the retina (presynaptic) and in the lateral geniculate nucleus (LGN) (postsynaptic) of an anesthetized rhesus monkey over a 10-second time window. Panel (c) shows a hippocampal place cell of a freely moving rat. Panel (d) plots the activity of a hippocampal neuron where some periodic spiking activity is present (horizontal dark band).*

of the brain. This similarity in the ISI plots of panels (a) and (b) is consistent with the results of Sincich et al. (2007) and Carandini et al. (2007) where the retinal ganglion cells and their paired lateral geniculate nucleus (LGN) neurons were studied. They, and the references therein, showed that not all the retinal spikes get transmitted from the retina to the cortex. In particular, Sincich et al. (2007) showed that it takes two retinal EPSPs to generate a spike in LGN. This is consistent with the more "sparse" pattern in panel (b), which represents the LGN data compared to panel (a), the retinal

data. Lastly, to see how the departure from time homogeneity is illustrated in the ISI plot, Figure 2.2 shows the ISI plots from a real dataset and a simulated homogeneous Poisson process. The two datasets share a common average intensity i.e., the average interspike interval is the same for the datasets in panels (a) and (b).

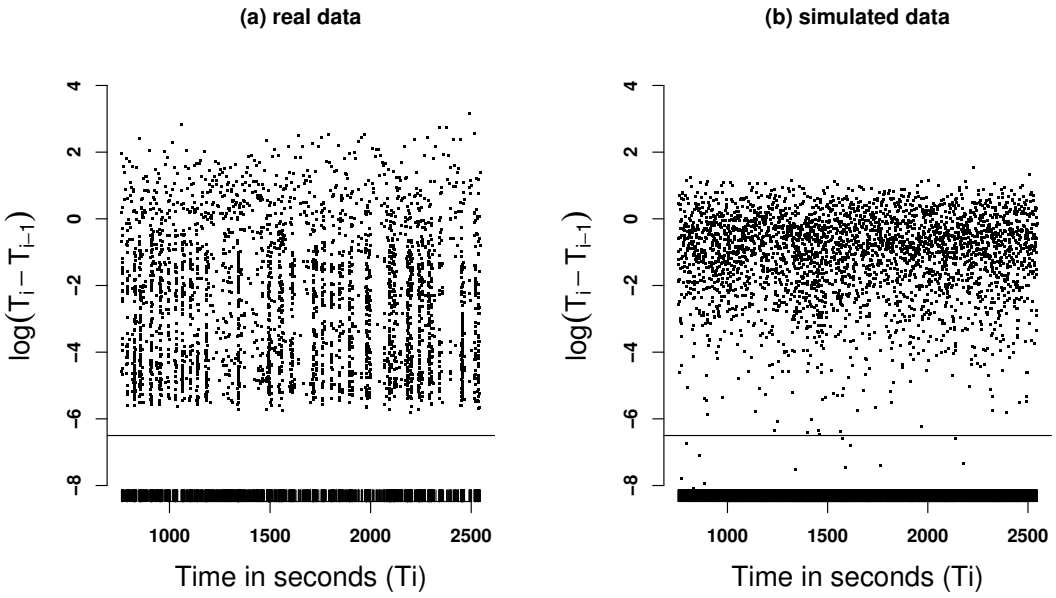


Figure 2.2: *The interspike interval plots: (a) real data - (b) simulated homogeneous Poisson process whose average intensity is the same as that of the real data in panel (a)*

While the real data shows both horizontal and vertical structures, the simulated data is more homogeneous. Furthermore, the minimum value on the y axis of the simulated data is much smaller than that of the real data, despite the average intensities of the two processes being the same. While in the real data [panel (a)] all of the interspike intervals are well above the refractory period limit (the black line), quite a few incidents in the simulated data [panel (b)] are below the refractory period.

On the basis of what has been said, interest lies in developing models for the intensity function of the neural spiking activity which are able to explain biological aspects of interest from different time scales, such as bursting, temporal inactivity, refractory period and periodic oscillations. Furthermore, these models should be able to model efficiently complex temporal relationships between a stimulus intensity and the observed spike trains.

2.4 Multiscale Modelling

Multiscale modelling in the neuroscience context, has been mostly applied to EEG data, particularly in reconstruction of the brain's electrical activity, Gavit et al. (2001), spatio-temporal analysis of hippocampal neurons, Sanchez et al. (2006), and identification of epileptic spikes, Calvagno et al. (2000); Indiradevi et al. (2008); Ganesan et al. (2010). There is, however, some literature on multiscale modelling of spike trains. The work of Nelson (2002) shows that variability in spike trains depends on the time scale, whose effect on the analysis of neural spike trains is studied in (Sanchez and Principe, 2007, pp.37-40). In another study, Suhail and Oweiss (2004) show that the performance of spike detection tests depends on time scale. Modelling experimental data from a motor control study, Kim et al. (2005) reported that multiscale modelling of the spike trains based on spike counts in time bins, enriches the representation of hand movement models. All of these studies motivate using models which can accommodate neural activity from different time scales, simultaneously. However, we haven't found any studies offering a tailored method to estimate the rate function of neural spike trains when there are obvious spiking patterns from multiple time scales.

2.4.1 Multiresolution Analysis for Likelihoods

We employ the multiscale penalized likelihood method introduced in Kolaczyk and Nowak (2004, 2005) to estimate the time-varying intensity function. Both references suggest recursive partitioning of the observation interval, followed by a pruning process, which results in merging some of the intervals based on the values of a penalized loglikelihood function. Before proceeding, we shall provide some details on this multiscale methodology.

Consider dividing the time interval $[0, T)$ by the set of splitting points $S = \{iT/N\}_{i=1}^{N-1}$. Starting with the original interval $[0, T)$, a *recursive partitioning (RP)* is produced by splitting only one of the intervals at a time at one of the unused splitting points in S until no splitting point remains, and thus the complete recursive partition (C-RP) $\mathscr{P}^*$ is obtained. As a special case, N can be a power of 2 i.e., $N = 2^J$ and at each step one of the intervals is divided exactly in half until the complete recursive partition is produced. This special case is called the *recursive dyadic partitioning (RDP)* and its associated C-RP is called the complete recursive dyadic partition (C-RDP). For

example, three possible complete recursive partitions of the interval $[0,1)$ are

$$\begin{aligned} P_1^* &= \Big\{ \{[0,1)\}\ \{[0,1/2),[1/2,1)\}\ \{[0,1/4),[1/4,1/2),[1/2,1)\} \\ &\quad \{[0,1/4),[1/4,1/2),[1/2,3/4),[3/4,1)\}\ \Big\}, \end{aligned}$$

$$\begin{aligned} P_2^* &= \Big\{ \{[0,1)\}\ \{[0,1/4),[1/4,1)\}\ \{[0,1/4),[1/4,3/4),[3/4,1)\} \\ &\quad \{[0,1/4),[1/4,1/2),[1/2,3/4),[3/4,1)\}\ \Big\}, \end{aligned}$$

$$\begin{aligned} P_3^* &= \Big\{ \{[0,1)\}\ \{[0,3/5),[3/5,1)\}\ \{[0,1/5),[1/5,3/5),[3/5,1)\} \\ &\quad \{[0,1/5),[1/5,2/5),[2/5,3/5),[3/5,1)\} \\ &\quad \{[0,1/5),[1/5,2/5),[2/5,3/5),[3/5,4/5),[4/5,1)\}\Big\}. \end{aligned}$$

Notice that P_1^*, P_2^* and P_3^* are each a collection of partitions. In the example above, P_1^* is a complete recursive dyadic partition, (C-RDP), P_2^* and P_3^* are non-dyadic complete recursive partitions (C-RP). Notice that the set of split points S for P_1^* and P_2^* is $\{i/(2^2)\}_{i=1}^3$, while that of P_3^* is $\{i/5\}_{i=1}^4$. During the recursive partitioning process (either dyadic or non-dyadic), the two intervals produced by splitting a so-called *parent interval* at each step are respectively called the *left* and the *right child*. For example in P_1^*, $[0,1/2)$ is the parent interval to $[0,1/4)$ (left child) and $[1/4,1/2)$ (right child). The complete recursive partitioning is also called a tree. Figure 2.3 displays P_1^* and P_2^* in the form of trees. Notice that the number of intervals at the bottom of the tree is usually associated with the term "resolution" as it determines the length of the sub-intervals at the bottom of the tree.

For computational efficiency, it is assumed here that $N = 2^J$, which does not necessarily imply the RDP (for example see P_2^* above). The value of J, which determines the value of N, is called the resolution. Therefore, for a given observation interval $[0,T)$, the finer the resolution, the more intervals at the bottom of the tree.

In another example, consider the two partitions $\mathscr{P}_1$ and $\mathscr{P}_2$ of the interval $[0,1)$ as follows;

$$\begin{aligned} \mathscr{P}_1 &= \{[0,1/2),[1/2,1)\}\ , \\ \mathscr{P}_2 &= \{[0,1/4),[1/4,1/2),[1/2,3/4),[3/4,1)\}\ , \end{aligned}$$

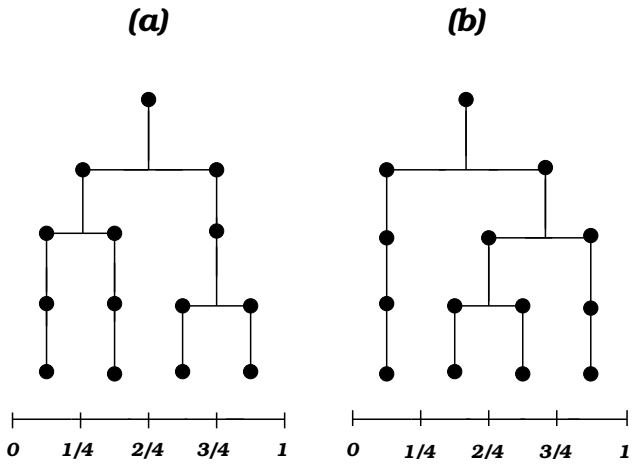


Figure 2.3: *The recursive partitioning of the interval* $[0, 1)$. *While (a) shows the dyadic tree, (b) represents a non-dyadic tree.*

where the elements of $\mathscr{P}_1$ are unions of the intervals in $\mathscr{P}_2$. We show this by $\mathscr{P}_1 \prec \mathscr{P}_2$, and say that $\mathscr{P}_2$ is a *refinement* of $\mathscr{P}_1$ (refinement that includes potential equivalence will be denoted using "$\preceq$"). Now, based on an arbitrary partition $\mathscr{P}$, define $\ell(\mathscr{P})$ to be the collection of all intervals I found in at least one partition $\mathscr{P}' \preccurlyeq \mathscr{P}$. Similarly, define $\ell_{NT}(\mathscr{P}) = \{I \in \ell(\mathscr{P}),\ I \notin \mathscr{P}\}$. For the partition P_1^* in the example above we have,

$$\begin{aligned} \ell(\mathscr{P}) &= \Big\{[0, \frac{1}{4}), [\frac{1}{4}, \frac{2}{4}), [\frac{2}{4}, \frac{3}{4}), [\frac{3}{4}, 1), [0, \frac{1}{2}), [\frac{1}{2}, 1), [0, 1)\Big\} \\ \ell_{NT}(\mathscr{P}) &= \Big\{[0, \frac{1}{2}), [\frac{1}{2}, 1), [0, 1)\Big\} \end{aligned}$$

Now, consider a stochastic process $X(t)$ observed on the N intervals $I_i \equiv [\frac{i}{N}, \frac{i+1}{N})$, $i = 0, 1, ..., N-1$. Let $X_0, X_1, ..., X_{N-1}$ and $\theta_0, \theta_1, ..., \theta_{N-1}$ denote, respectively, the number of observations and the mean function of the process associated with the N subintervals. Kolaczyk and Nowak (2004) introduced the following set of four conditions based on which a so-called multiscale factorization of the likelihood function is constructed.

1. "**Hierarchy of recursive partitions.** A hierarchy of recursively defined partitions
$$... \prec \mathscr{P}_{\ell-1} \prec \mathscr{P}_\ell \prec \mathscr{P}_{\ell+1} \prec ... \tag{2.4.1}$$
beginning with $[0, 1)$ and ending with a complete recursive partition C-RP $\mathscr{P}^* = \{I_i\}_{i=0}^{N-1}$."

2. "**Independence within $\mathscr{P}^*$.** The components of $\mathbf{X} = (X_0, X_1, ..., X_{N-1})$ are statistically independent and the components of $\boldsymbol{\theta}$ are L-independent with respect

to the likelihood of $\mathbf{X}$, that is,

$$p(\mathbf{X}|\boldsymbol{\theta}) = \prod_{i=0}^{N-1} p(X_i|\theta_i). \tag{2.4.2}$$

where $\theta_i \in \Theta_i,\ i = 0, 1, ..., N-1$ and $\boldsymbol{\theta} \in \Theta = \Theta_0 \times \Theta_1 \times \cdots \times \Theta_{N-1}$. Furthermore, the p.d.f. of each X_i is a member of some common parametric family $\mathscr{F} \equiv \{p(.|\theta);\ \theta \in \Theta \subseteq \mathbb{R}\}$."

3. "**Reproducibility between partitions.** The family $\mathscr{F}$ is reproducible in θ, in the sense that, for all $I \in \ell(\mathscr{P}^*)$ and $\forall \boldsymbol{\theta} \in \Theta^{N_I}$, the p.d.f of $X_I \equiv \sum_{i/N\in I} X_i$ is $p(X_I|\theta_I) \in \mathscr{F}$, where $\theta_I \equiv \sum_{i/N\in I} \theta_i$."

   Although this condition results in computational convenience, it is too limiting. For example, if θ_i's are to be approximated with linear functions rather than piecewise constant functions, this condition will not be satisfied.

4. "***Decoupling* of parameters with partitions (i.e. cuts).** For any $X_i \sim p(.|\theta_i) \in \mathscr{F}$, $i \in \{i_1, i_2\}$, there exists some reparameterization $(\theta_{i_1}, \theta_{i_2}) \to (\theta, w)$ such that

$$p(X_{i_1}, X_{i_2}|\theta_{i_1}, \theta_{i_2}) = p(X|\theta)p(X_{i_1}|X, w), \tag{2.4.3}$$

   where $X \equiv X_{i_1} + X_{i_2}$ and $\theta \equiv \theta_{i_1} + \theta_{i_2}$. That is, the sum X is a cut for (X_{i_1}, X_{i_2}). Hence, the marginal inference of the multiscale parameters is possible."

   Based on a family of distributions $\mathscr{F}$ the statistic T is called a cut if there exists a parametrization $\{\mathscr{F}_\omega :\ \omega \in \Omega\}$ and partition (ω_1, ω_2) of ω such that ω_1 and ω_2 are L-independent and $p_\omega(x) = p_{\omega_1}(t) \times p_{\omega_2}(x|T = t)$. For the details on cuts refer to Barndorff-Nielsen (1976, 1978); Bar-Lev and Pommeret (2003).

Theorem 2.1 *[Kolaczyk and Nowak (2004)] Under the conditions 1-4 there exists a factorization of the form*

$$p(\mathbf{X}|\boldsymbol{\theta}) = p(X_{I_{00}}|\theta_{I_{00}}) \prod_{I\in\ell_{NT}(\mathscr{P}^*)} p(X_{ch(I),l}|X_I, w_I) \tag{2.4.4}$$

with respect to some reparameterization $[\theta_{I_{00}}, w]$ *of* $\boldsymbol{\theta}$*, for* $I_{00} \equiv [0, 1)$ *and* $\theta_{I_{00}} \equiv \sum_{i=0}^{N-1} \theta_i$.

The conditions 1-4, which allow for a likelihood multiresolution analysis (likelihood MRA), are, essentially, similar to the conditions of regular multiresolution analysis, under which the wavelet decomposition of an L^2 function is derived. Table 2.1 (Kolaczyk and Nowak, 2003, p.252) shows a comparison of the conditions for likelihood MRA and those of the wavelet MRA.

Table 2.1: Comparison of wavelet and likelihood MRA

Wavelet MRA	**Likelihood MRA**
Hierarchy of nested subsets	Hierarchy of recursive partitions
Orthonormal basis within V_0	Independence within $\mathscr{P}^*$
Scalability between subspaces	Reproducibility between partitions
Translation within subspaces	*Decoupling* of parameters with partitions (i.e. cuts)

In summary, the wavelet expansion of an L^2 function based on the wavelet MRA translates to the multiscale factorization of the likelihood function in Equation (2.4.4). This factorization allows for the statistical inference at different scales independent of other scales. It also allows for marginal inference on the parameters ω_I.

2.4.2 A multiresolution probability model for counts

Similar to Brown et al. (2002); Sanger (2002); Dayan and Abbott (2001); Lee et al. (2010); Kim and Basso (2010), in this chapter, we model neural spike trains as realizations of an inhomogeneous Poisson process. Therefore, it is interesting to study the application of Theorem 2.1 on the Poisson family. Kolaczyk and Nowak (2004) derived the multiscale probability model for such a distributional family. Consider the function $\theta \in \Theta$, where $\theta(t) \in [l, u]$, for all $t \in [0, 1]$. Let $0 < l < u$, and define $\theta_i = N \int_{I_i} \theta(t)dt$ to be the average of θ over I_i. Furthermore, let $X_i \sim Pois(\theta_i)$. Then the components in the likelihood factorization (2.4.4) take the form

$$\begin{aligned} X_{ch(I),l} \mid X_I, w_I &\sim Bin(X_I\,,\, w_I)\,, \\ X_{I_{00}} \mid \theta_{I_{00}} &\sim Pois(\theta_{I_{00}}), \end{aligned}$$

where $w_I = \theta_{ch(I),l}/\theta_I$.

2.4.3 Multiscale penalized likelihood

In the recursive partitioning setup (dyadic or non-dyadic), we denote each interval (node) by $I_{jk},\ j = 0, ..., N,\ k = 0, ..., j-1$, where j shows the level and k represents the location of the interval in that level. Let X_{jk} be the random variable representing the number of spikes in that interval, where $E(X_{jk}) = c_{jk}$. Also, let X_0 represents the number of spikes in the whole observation window $[0, T)$. Let $\widehat{c}(t)$ be the multiscale estimate of the intensity function of the underlying Poisson process. Given N, the resolution of the complete recursive tree, we have $\widehat{c}_{Nk} = \int_{kT/N}^{(k+1)T/N} \widehat{c}(t)\, dt$. In this thesis we assume that $\widehat{c}(t)$ is a piecewise constant function, which implies that $\widehat{c}(t) = \widehat{c}_{Nk}/(T/N)\,,\ \ kT/N < t \leq (k+1)T/N$.

To control the complexity of the estimated intensity function, we use the penalized likelihood method. The penalized likelihood function is shown by $\ell(c) - \lambda \text{pen}(c)$, where $\ell(c)$ is the Poisson loglikelihood function $[X_{jk} \sim \text{Pois}(c_{jk})]$, λ is the penalty factor, and $\text{pen}(c)$ is a penalty function, penalizes the loglikelihood function for number of parameters $c_{jk},\ j = 0, ..., N,\ k = 0, ..., j-1$. The two child intervals I_{jk}^{ℓ} and I_{jk}^{r} are merged if the penalized loglikelihood of the parent interval I_{jk} is larger than sum of those at the two child intervals. The penalty function $\text{pen}(c)$ used in this thesis is simply the number of parameters at each level, see Kolaczyk and Nowak (2005). In summary, if the intensity values are similar enough across the two child intervals (consecutive bins), the two bins are merged, and the complexity of the estimated intensity is reduced. For details on penalized likelihood methods refer to Eggermont and LaRiccia (2001), where density/intensity function estimation based on penalized likelihoods is discussed. Note that, while increasing N (or resolution) results in more volatile fits, increasing λ (penalty factor) produces "smoother" fits and vice versa.

Pruning the recursive tree merges some of the time bins, which results in bins B_i of different sizes ($[0, T) = \cup_i B_i$). One of the neuroscientific gains of allowing for multiple time scales is that it identifies subintervals of the observation window $[0, T)$, during which the spiking rate is constant.

2.4.4 Choosing tuning parameters via cross-validation

The tuning parameters of the multiscale model are N and λ. The smaller the N or larger λ, the smoother the fit, which shows that the tuning pair (N, λ) can quantitatively control for over-fitting. Therefore, proper choice of these parameters is particularly important to avoid over/under fitting. Since it is assumed that $N = 2^J$, we can think of either N or J as the resolution parameter. Given the data, let I_{min} and T

show the length of the shortest interspike interval and the length of the observation window, respectively. It is therefore natural to assume that $T/N > I_{min}$. This inequality imposes a practical upper bound on N, or equivalently on J. Another option, which is more conservative, is suggested in (Sanchez and Principe, 2007, p.37), where it is assumed that $T/N > 1$ millisecond, setting the highest resolution at the refractory period of neural firing.

Let us assume that the data consists of m trials i.e., m spike trains from the same neuron whose unknown intensity function is $\theta(t)$. Let $\widehat{c}_i(t)$ denote the multiscale estimate based on the i^{th} spike train, consisting of n_i spikes, $i = 1, 2, ..., m$, which is clearly a function of N and λ. We employ a leave-one-trial-out cross-validation based on the integrated squared error loss (ISE). At each step of the m iterations of the cross-validation, one of the m trials is omitted. Equation (2.4.5) provides the estimated pair (N^*, λ^*) via cross-validation based on ISE. The derivation of this equation is provided in Section 2.7.

$$(N^*, \lambda^*)_{ISE} = \arg\min_{N,\lambda} \left\{ \int_0^T \Big(\sum_{i=1}^{m} \frac{\widehat{c}_i(t)}{m\, n_i} \Big)^2 dt - \frac{2}{m} \sum_{i=1}^{m} \frac{\sum_{\ell=1}^{n_i} \sum_{j \neq i} \widehat{c}_j(t_{i\ell})/n_j}{n_i(m-1)} \right\}, \qquad (2.4.5)$$

The time of the ℓ^{th} spike of the i^{th} trial is represented by $t_{i\ell}$. Notice that this is the cross-validation method derived from the kernel estimation method, see (Givens and Hoeting, 2005, p.285). Our approach is not kernel-based and this method could have provided biased estimates of the intensity function. However, our simulation study shows that any potential bias in the multiscale estimate of the intensity function is small (see Figure 2.4). We have also studied the robustness of the tuning parameters in Equation (2.4.5). Based on a simulation study, Figure 2.4 shows that the "optimum" parameters $(N^*, \lambda^*)_{ISE}$ calculated from (2.4.5) can reproduce the true intensity function. Notice that the bootstrap confidence bands in Figure 2.4 are based on asymptotic results, therefore, as the sample size increases, the coverage of the confidence bands improve. The effect of the sample size (100 versus 1000 realizations) is plotted in the bottom panel of Figure 2.4.

If the stimulus signal is available, and interest lies in understanding the relationship between the signal and the intensity function, one could optimize the tuning parameters (N, λ) so that most of the variability in the stimulus signal is explained by the intensity function. To do so, any loss function could be used to measure the similarity between the two functions. To perform such cross-validation, the *latency* to the stimulus on-set should be estimated, which is discussed briefly in our real data analysis in Section

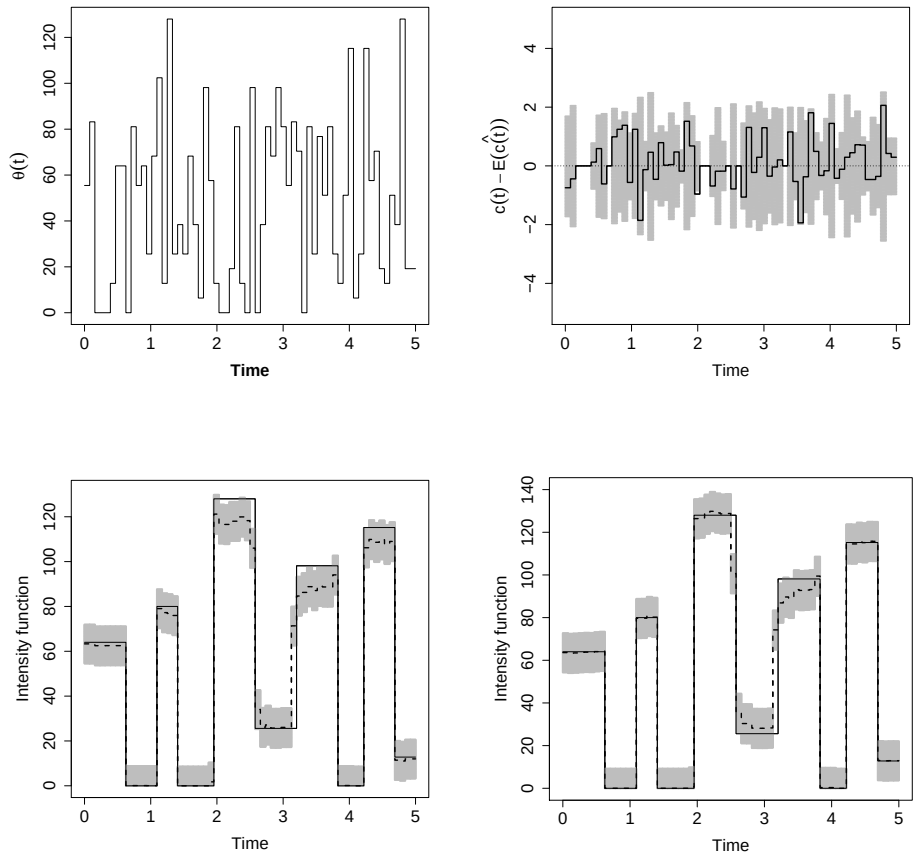


Figure 2.4: *The performance of the cross validation method. Top left: The true intensity function. Top right: The bias of the estimated intensity function along with a 95% confidence band calculated based on 1000 simulations (gray area). Bottom left: The true intensity function (solid line) and its estimate (dashed line) based on 100 simulations of the inhomogeneous Poisson process along with the 95% confidence band (gray area). Bottom right: The true intensity function (solid line) and its estimate (dashed line) based on 1000 simulations of the inhomogeneous Poisson process along with the 95% confidence band (gray area)*

3.2. Notice that forcing the similarity between the stimulus signal and the spiking rate needs biological justification. Figure 5.5a shows situations where the spiking rate changes in the same or in the opposite direction relative to changes in the stimulus signal.

2.5 Modelling periodicity and brain rhythms

One of the properties of neural activity is the periodicity or the rhythms of the brain. As mentioned in the previous chapter, neural oscillatory activity, which is an inevitable property of the brain (Buzsáki, 2006, pp.111-119), is categorized according to the frequency range; δ-rhythm (2-4Hz), Walker (1999), θ-rhythm (4-8Hz), Miller (1991); Malhotra et al. (2012), α-rhythm (8-13Hz), Windhorst and Johansson (1999), β-rhythm (13-30Hz), Lopes da Silva (1991) and γ-rhythm (more than 30Hz), Freeman (1992). Some of these rhythms can be simultaneously present in a given area of the brain. Simultaneous presence of multiple frequencies in a vision experiment on monkeys is reported in Bressler et al. (1993). According to Fischer et al. (2002), θ and γ rhythms characterize the hippocampal activity *in vivo*. Furthermore, task-specific cognitive performance is related to the interactions between different brain rhythms, Tort et al. (2008, 2009); Axmacher et al. (2010). To integrate cross-frequency information, a phase-phase coupling of *gamma* and *theta* oscillations in the CA1 region of rat hippocampus is discussed in Belluscio et al. (2012). For a review on cross-frequency coupling refer to Jensen and Colgin (2007). For more details of the brain rhythms, refer to Buzsáki (2006). Besides brain rhythms, significant frequencies in the temporal power spectrum of the stimulus signal can also carry over to the spike trains. An example of such a case is discussed in Chapter 3.

The literature has addressed the analysis of periodicity in the point process framework. Bartlett (1963) discussed spectral analysis for the univariate point processes and introduced the periodogram of the point process data. He also showed that the asymptotic properties of this periodogram are similar to those of the conventional periodogram of continuous time series. He has discussed some details of this work in his book, (Bartlett, 1978, pp.342-352), whose extension to bivariate case is addressed in Bartlett (1964). Along the same lines, Lewis (1970, 1972) proposed a periodic intensity function for an inhomogeneous Poisson process with known frequency, and Vere-Jones (1982) extended the works of Lewis (1970, 1972) to the case where it is only known that the frequency is in a specific range. Vere-Jones (1982) also studied the asymptotic properties of the frequency estimate i.e., the value corresponding to the maximum of Bartlett's periodogram. Based on spike train data, Rigas and Tsitsis (1996) performed a bivariate spectral analysis on gamma and alpha motoneurons to show that the presence of an alpha motoneuron reduces the effect of a gamma motoneuron on the muscle spindle. In extending the works of Lewis (1970, 1972) and Vere-Jones (1982), Shao and Lii (2011) proposed almost periodic intensity functions for inhomogeneous Poisson

processes as the sum of sinusoidal functions plus a baseline

$$\theta(t) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k) + B, \tag{2.5.6}$$

where A_k, B, ω_k and ϕ_k are unknown parameters satisfying some basic conditions so that $\theta(t)$ is an intensity function. They provided consistent estimates of the frequencies, phases, and amplitudes and discussed their asymptotic statistical properties in detail. In a different context, on γ-ray pulsar detection, Bickel et al. (2007, 2008) proposed the following model for the time-varying intensity function of an inhomogeneous Poisson process;

$$\theta(t) = \mu s(t) \left[(1-\theta) + \theta \nu_\tau(ft)\right], \tag{2.5.7}$$

where $\nu_\tau(t)$ is a periodic function with the initial phase τ. While f, τ and μ are unknown parameters, the function $s(t)$ is known, and accounts for the sensitivity of the recording device at time t. They developed a score test to investigate the significance of the periodic component in this model. We are interested in combining a piecewise constant function $c(t)$ over our partition with some periodic terms, which accommodate brain rhythms and/or periodicity in the stimulus signal. Combining the model of Shao and Lii (2011) [multiple periodic terms in Equation (2.5.6)] with the model of Bickel et al. (2007, 2008) to include the effect of the multiscale function $c(t)$, Equation (2.5.7), we propose the following multiplicative $\theta_m(t)$ and additive $\theta_a(t)$ intensity functions for neural spike trains;

$$\theta_m(t) = c(t)\Big\{(1 - \sum_{k=1}^{K} \eta_k) + \sum_{k=1}^{K} \eta_k \, \nu_{\gamma_k}\big(f_k t + \omega_k^{(0)}\big)\Big\}, \tag{2.5.8}$$

$$\theta_a(t) = \left(1 - \sum_{k=1}^{K} \eta_k\right) c(t) + \sum_{k=1}^{K} \eta_k \nu_{\gamma_k}\big(f_k t + \omega_k^{(0)}\big), \tag{2.5.9}$$

where $\boldsymbol{\eta} = (\eta_1, ..., \eta_K)$, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)$, $\boldsymbol{f} = (f_1, ..., f_K)$ and $\boldsymbol{\omega^{(0)}} = (\omega_1^{(0)}, ..., \omega_K^{(0)})$ are vector parameters of length K. The restrictions $\eta_k \geq 0$ for $k = 1, ..., K$ and $\sum_k \eta_k \leq 1$ guarantee the positivity of the intensity function. The periodic functions ν_{γ_k}, $k = 1, ..., K$ are defined by

$$\nu_{\gamma_k}(x) = \gamma_k \left[1 + \cos(2\pi\, x)\right], \qquad \text{for} \quad \omega_k^{(0)} < x < f_k\, T + \omega_k^{(0)}.$$

In these equations $f_k > 0$ and $-\frac{1}{4} < \omega_k^{(0)} < \frac{3}{4}$, $k = 1, .., K$ are, respectively, frequency values (in Hz) and initial phases, Shao and Lii (2011). Notice that while the non-negative parameters η_k, $k = 1, ..., K$, quantify the contribution of the periodic components relative to the baseline multiscale function $c(t)$, the parameters $\gamma_k \geq 0$,

$k = 1, ..., K$ are the amplitudes of these periodic components. The derivations of the loglikelihoods based on these intensity functions are available in Section 2.7.

The smoothed periodogram has been employed to estimate the frequencies $\boldsymbol{f}$ (Brillinger, 2001, pp.131-142). We have noticed, through simulation studies, that the smoothed periodogram significantly outperforms the raw periodogram in terms of detecting the correct frequency when $c(t)$ is a piecewise constant function. We have generated 1000 realizations of a Poisson process with the intensity function

$$\theta(t) = c(t)\Big(2 + \cos(4\sqrt{2}\,\pi\, t)\Big) \quad 0 \leq t < 20,$$

where $c(t)$ is a piecewise constant function displayed in Figure 2.5.

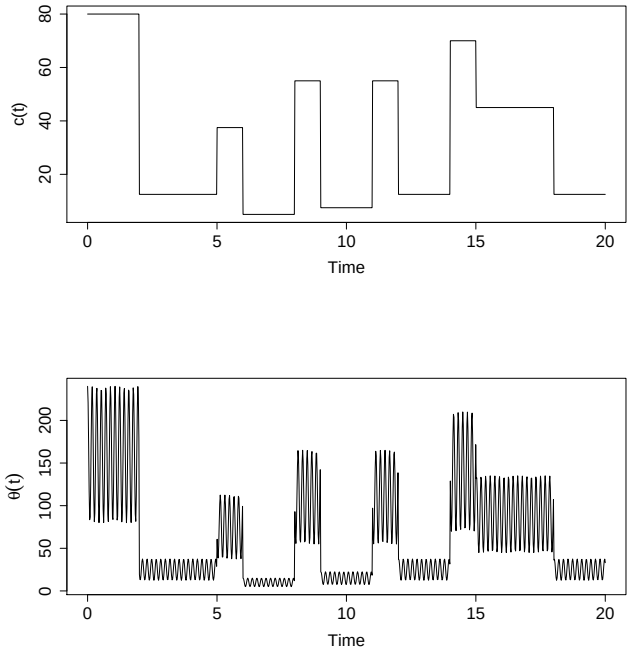


Figure 2.5: *The intensity function. upper panel: the multiscale estimate* $c(t)$ *- lower panel: the intensity function,* $\theta(t) = c(t)\Big(2 + \cos(8\sqrt{2}\pi t)\Big)$.

Utilizing the classic periodogram, we have noted that more than 30% of the time, the difference between the optimal frequency and the true frequency is more than 0.3Hz. However, this error rate using the smoothed periodogram drops to under 3%. This may be because spectral leakage (Arrillaqa and Watson, 2003, pp.39-41) affects the two periodograms differently. It may also be the effect of the piecewise constant function $c(t)$

on the periodic behaviour of the process. For details about smoothed periodograms refer to (Brillinger, 2001, 131-142) and the references therein.

To estimate the frequencies $\boldsymbol{f}$, based on an initial search of the smoothed periodogram, K frequencies corresponding to the K largest peaks of the smoothed periodogram are extracted and then a more refined search in the neighbourhood of each frequency is performed. This method is suggested in Shao and Lii (2011), however, we substitute the raw periodogram with the smoothed periodogram in thier method, which improves the frequency estimation with some computational cost. Since the initial phases $\boldsymbol{\omega^{(0)}}$ are functionally dependent on the frequency values $\boldsymbol{f}$, Shao and Lii (2011), we treat the K frequencies, which result in the K highest values of the smoothed periodogram, and their associated initial phases $\boldsymbol{\omega^{(0)}}$, as plug-in estimates. We then numerically maximize the $2K$-parameter loglikelihood function for $\boldsymbol{\eta} = (\eta_1, ..., \eta_K)$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)$ using the Nelder-Mead algorithm Nelder and Mead (1965); Lagarias et al. (1998). In our data analyses, we have noticed that having the plug-in estimates of $\boldsymbol{f}$ and $\boldsymbol{\omega^{(0)}}$, the algorithm converges to the same solution for $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ regardless of the choice of the initial parameter values. It is noteworthy that the function $c(t)$ and the frequencies $\boldsymbol{f}$ are estimated separately. Based on a set of spike trains, the function $c(t)$ is estimated by the penalized likelihood method discussed in Section 2.4. The frequencies are estimated separately using the smoothed periodogram on the spike train data. These estimates are then put together in the two models (2.5.8) and (2.5.9).

According to Lii Shao and Lii (2011), choosing the number of significant frequencies in general is still an open question, but model selection techniques can be used to select K. We initially choose the frequencies which appear in at least half of the trials. Then, we fit all models with possible subsets of these frequencies and choose the final model according to a model selection criterion such as the Akaike information criterion (AIC), Akaike (1974), corrected Akaike information criterion (AICc), Hurvich and Tsai (1989), or Bayesian Information Criterion, Schwarz (1978). To check the performance of these model selection criteria, we have performed a simulation study. One hundred trials were generated from a multiplicative intensity function with $K = 3$. The parameter values used in this simulation are $f = (2.8, 5, 6.5)$, $\boldsymbol{\omega^{(0)}} = (0.4, 0.5, -0.15)$, $\boldsymbol{\eta} = (0.1, 0.4, 0.2)$ and $\boldsymbol{\gamma} = (0.5, 0.7, 1)$. Table 2.2 shows the results of the simulation study.

Although the selection criteria unanimously suggest a multiplicative model with $K = 2$ periodic components (as opposed to $K = 3$), we have noticed that the fit of this model relative to the true intensity is remarkably good. The frequency value which has not

Table 2.2: *Average values of different model selection criteria. The true model has* $K = 3$ *periodic components.*

	K=1		K=2		K=3		K=4	
	Mult.	**Addit.**	**Mult.**	**Addit.**	**Mult.**	**Addit.**	**Mult.**	**Addit.**
AIC	-1957.1	-1944.9	**-1958.1**	-1951.9	-1950.3	-1936.7	-1942.6	-1927.7
AICc	-1943.6	-1931.4	**-1947.2**	-1941.1	-1933.9	-1920.2	-1922.8	-1907.9
BIC	-1809.2	-1797.0	**-1824.9**	-1818.8	-1787.6	-1773.9	-1765.1	-1750.2

been picked by any of the model selection criteria is $f = 2.8Hz$. Notice that the contribution of each periodic term in Equations (2.5.8) and (2.5.9) depends on the parameter values $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$. In this simulation, the frequency 2.8Hz is associated with the smallest values in both parameters. Therefore, the contribution of this periodic term is not as big as those of the other two periodic components (associated with 5 and 6.5Hz frequencies). Not surprisingly, 2.8Hz is the frequency which has not been identified in the selected model with $K = 2$ periodic components. Figure 2.6 plots the true and the estimated intensity functions. The true intensity function (black solid curve) has $K = 3$ periodic components with $\boldsymbol{f} = (2.8, 5, 6.5)$ and the estimated intensity (dashed red curve), selected with the model selection criteria (AIC, AICc, BIC), has $K = 2$ periodic components with $\widehat{\boldsymbol{f}} = (5, 6.4)$.

2.6 Discussion

In this chapter we proposed a general and rich family of intensity functions for the neural spiking activity in an inhomogeneous Poisson process framework. Multiscale additive and multiplicative models with periodic components were introduced. We also employed the powerful ISI plot, which can visualize biological phenomena from different time scales. It is important to emphasize that the multiscale framework laid out in this chapter is computationally fast. The computational efficiency of our model is particularly noteworthy because in many situations spike trains are collected over a long period of time, which makes the intensity function estimation problem a computationally intensive task. Our models are fast enough that computational issues are not the main concern in their implementation. The following are some details and discussions about several observations we have had about the topics discussed in this chapter.

The first observation is about the difference between the recursive partitioning and the recursive dyadic partitioning estimators. Let $\widehat{c}_{RP}(t)$ represent the multiscale estimate

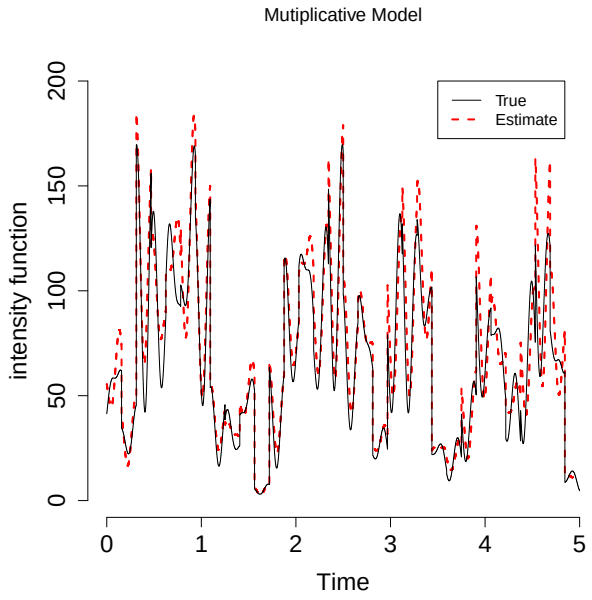


Figure 2.6: *Comparison between the estimated (K=2) and the true (K=3) intensity function.*

of the intensity function based on the recursive partitioning (RP), and $\widehat{c}_{RDP}(t)$ that of the recursive dyadic partitioning (RDP). Kolaczyk and Nowak (2004) have shown that $\widehat{c}_{RDP}(t)$ and $\widehat{c}_{RP}(t)$ can be computed using $O(N)$ and $O(N^3)$ penalized likelihood comparisons, respectively. Here, N is the number of intervals at the bottom of the recursive partitioning tree. Clearly, for large N, the difference between the computation time of $\widehat{c}_{RP}$ and $\widehat{c}_{RDP}$ becomes an important issue. We noticed that the two methods are also different in terms of the *"windowing effect."* To explain this effect, consider the 5 intervals shown in Figure 2.7. Starting from interval #1, at each level, the observation window has been slightly shifted to the right. The *"windowing effect"* of a given method is then defined to be the difference between the estimates of the intensity function produced by the method, in the intersection of these intervals (gray area). Clearly, we would like the windowing effect to be as as small as possible. The effect of partitioning on statistical inference has been also addressed in a different context by Ferguson (1974) and Paddock et al. (2003).

We have noticed that $\widehat{c}_{RP}(t)$ has much lower windowing effect. In other words, $\widehat{c}_{RP}(t)$ is more stable during the shaded area across the different observation intervals. This is not surprising since the recursive dyadic partitioning only allows for dyadic splits of the hierarchical intervals. This imposes limitations on the structure of the recur-

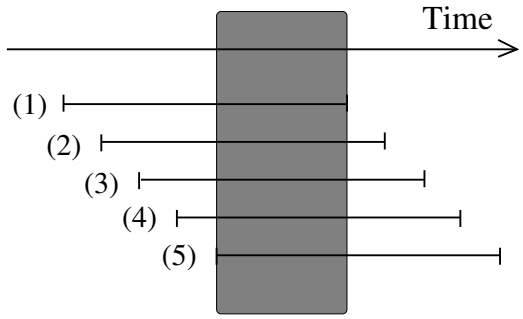


Figure 2.7: *The intersection of consecutive intervals, during which the windowing effect can be noticed.*

sive partitioning, and makes the corresponding estimator $\widehat{c}_{RDP}(t)$, quite sensitive to small changes in the end-points of the observation interval $[0, T)$. However, $\widehat{c}_{RP}(t)$ is estimated by optimizing over a "forest of trees" rather than over the class of dyadic trees, Kolaczyk and Nowak (2005). The almost negligible windowing effect of the RP estimator motivated us not to employ the computationally efficient RDP estimator. Having said that, it is noteworthy that despite the limited dyadic structure, for a fixed observation window, $\widehat{c}_{RDP}(t)$ is a reasonable fit and is significantly faster than $\widehat{c}_{RP}(t)$ to compute, particularly, as N increases. For risk optimality comparisons between $\widehat{c}_{RP}$ and $\widehat{c}_{RDP}$ refer to Kolaczyk and Nowak (2004), where it has been shown that under mild conditions, and based on a squared Hellinger loss function, the risk of the estimators $\widehat{c}_{RP}$ and $\widehat{c}_{RDP}$ are bounded above by $O((\log N/N)^{2/3})$ and $O((\log^2 N/N)^{2/3})$, respectively.

Last, but not least, we would like to mention that the multiscale, multiple frequency methodology laid out in this manuscript (inhomogeneous Poisson process) can be extended to the non-Poisson cases. The IMI model of Kass and Ventura (2001) and the model of Pillow et al. (2008) are some examples of non-Poisson models for neural spike trains. To the best of our knowledge, none of the common non-Poisson models are in the multiscale, multiple frequency framework. The intensity function of non-Poisson spike trains can still be estimated via multiscale methods with appropriate modifications, however, the optimality results of the multiscale estimator provided by Kolaczyk and Nowak (2004, 2005) are only valid under a specific family of distributions. In Chapter 4 and Chapter 5 a biologically justifiable model for neural spike trains (called Skellam Process with Resetting) is developed, in which the parameters are estimated based on the multiscale method laid out in this paper.

2.7 Cross-validation and likelihood derivations

2.7.1 Derivation of the cross-validation method

Here we derive the cross-validation formula introduced in Equation (2.4.5). The estimate of the multiscale intensity based on the i^{th} replication, $\widehat{c}_i(t)$, has been scaled so that $\int_0^T \widehat{c}_i(t)\,dt = n_i$. Defining $f(t) = c(t)/\int_0^T c(t)\,dt$, $\widehat{f}_i(t) = \widehat{c}_i(t)/n_i$ the problem becomes a density function estimation. Let

$$\begin{aligned} \widehat{f}^{-i}(t) &= \sum_{j\neq i} \frac{\widehat{f}_j(t)}{m-1} \\ &= \sum_{j\neq i} \frac{\widehat{c}_j(t)}{n_j(m-1)}\,, \end{aligned} \tag{2.7.10}$$

and

$$\begin{aligned} \widehat{f}(t) &= \sum_{i=1}^{m} \frac{\widehat{f}_i(t)}{m} \\ &= \sum_{i=1}^{m} \frac{\widehat{c}_i(t)}{m\,n_i}\,. \end{aligned} \tag{2.7.11}$$

Following similar derivation of the cross-validation method in (Givens and Hoeting, 2005, p.285), we can write the integrated square error as

$$\begin{aligned} ISE(J,\lambda) &= \int_0^T \Big[f(t) - \widehat{f}(t)\Big]^2\,dt \\ &= \int_0^T [\widehat{f}(t)]^2\,dt - 2E\Big(\widehat{f}(t)\Big) + \int_0^T [f(t)]^2\,dt\,. \end{aligned}$$

The last term in the equation above is constant. Although we employ a multiscale estimation technique, we use the conventional estimate $\sum_{i=1}^{m} \widehat{f}^{-i}(t)/n$ for the expectation in the second term. This is adapted from the conventional cross-validation method for the kernel density estimation. Thus, minimizing

$$\begin{aligned} CV(J,\lambda) &= \int_0^T [\widehat{f}(t)]^2\,dt - 2\widehat{E\Big(\widehat{f}(t)\Big)} \\ &= \int_0^T [\widehat{f}(t)]^2\,dt - \frac{2}{m}\sum_{i=1}^{m} \frac{\sum_{\ell=1}^{n_i} \widehat{f}^{-i}(t_{i\ell})}{n_i} \end{aligned}$$

should be a reasonable choice for the parameters (J,λ). Substituting $\widehat{f}(t)$ and $\widehat{f}^{-i}(t)$ from Equations (2.7.10) and (2.7.11), the proof is complete.

2.7.2 Loglikelihood derivation for the multiplicative model

Recall that

$$\theta(t) \;=\; c(t)\Big\{(1-\sum_{k=1}^{K}\eta_k)+\sum_{k=1}^{K}\eta_k\nu_{\gamma_k}\big(f_k\times t+\omega_k^{(0)}\big)\Big\}$$

is the proposed intensity function. The loglikelihood function based on the spike train $\{t_i, i=1,...,n\}$ is

$$\ell(\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{f},\boldsymbol{\omega^{(0)}};n)=\sum_{i=1}^{n}\log\Big(\theta(t_i)\Big)-\int_0^T\theta(t)\,dt\,.$$

Notice that $c(t)$ is a piecewise constant function over the observation window $(0,T]=\cup_j(D_j,D_{j+1}]$, where $D_j,\ j=1,2,...,N$ are the breakpoints of the subintervals at the bottom of the recursive tree and $\Delta D_j=D_{j+1}-D_j$. Therefore, $\int_0^T c(t)\,dt=\sum_j c_j\Delta D_j$ where $c_j=c(t),\ t\in(D_j,D_{j+1}]$. We have

$$\begin{aligned}
\ell(\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{f},\boldsymbol{\omega^{(0)}};n) &= \sum_{i=1}^{n}\log\Big(\theta(t_i)\Big)-\int_0^T\theta(t)\,dt\\
&= \sum_{i=1}^{n}\log c(t_i)+\underbrace{\sum_{i=1}^{n}\log\left\{1-\sum_{k=1}^{K}\eta_k+\sum_{k=1}^{K}\eta_k\nu_{\gamma_k}\big(f_kt+\omega_k^{(0)}\big)\right\}}_{A}+\\
&\quad -\int_0^T c(t)\Big[1-\sum_{k=1}^{K}\eta_k+\sum_{k=1}^{K}\eta_k\gamma_k+\sum_{k=1}^{K}\eta_k\gamma_k\cos\Big(2\pi(f_kt+\omega_k^{(0)})\Big)\Big]\,dt\\
&= \sum_{i=1}^{n}\log c(t_i)-\int_0^T c(t)+A-\sum_{j=1}^{N}\int_{D_j}^{D_{j+1}}c_j\left[-\sum_{k=1}^{K}\eta_k+\right.\\
&\quad \left.+\sum_{k=1}^{K}\eta_k\gamma_k+\sum_{k=1}^{K}\eta_k\gamma_k\cos\Big(2\pi(f_kt+\omega_k^{(0)})\Big)\right]dt\\
&= \sum_{i=1}^{n}\log c(t_i)-\sum_{j=1}^{N}c_j\Delta D_j+A-\sum_{j=1}^{N}c_j\Big(\sum_{k=1}^{K}\eta_k(\gamma_k-1)\Big)\Delta D_j+\\
&\quad -\sum_{j=1}^{N}\sum_{k=1}^{K}c_j\frac{\eta_k\gamma_k}{2\pi f_k}\left[\sin\Big(2\pi(f_kD_{j+1}+\omega_k^{(0)})\Big)-\sin\Big(2\pi(f_kD_j+\omega_k^{(0)})\Big)\right]\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\ell(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{f}, \boldsymbol{\omega^{(0)}}; n) &= \sum_{i=1}^{n} \log c(t_i) - \sum_{j=1}^{N} c_j \Delta D_j \\
&\quad + \sum_{i=1}^{N} \log \left\{ 1 - \sum_{k=1}^{K} \eta_k + \sum_{k=1}^{K} \eta_k \gamma_k + \sum_{k=1}^{K} \eta_k \gamma_k \cos\left(2\pi(f_k t_i + \omega_k^{(0)})\right) \right\} \\
&\quad - \sum_{j=1}^{N} \sum_{k=1}^{K} \eta_k (\gamma_k - 1) c_j \Delta D_j \\
&\quad - \sum_{j=1}^{N} \sum_{k=1}^{K} c_j \frac{\eta_k \gamma_k}{2\pi f_k} \left[\sin\left(2\pi(f_k D_{j+1} + \omega_k^{(0)})\right) - \sin\left(2\pi(f_k D_j + \omega_k^{(0)})\right) \right] .
\end{aligned}$$

2.7.3 Loglikelihood derivation for the additive model

Similar to the multiplicative case, for $0 < t \leq T$ we can write,

$$\theta(t) = \left(1 - \sum_{k=1}^{K} \eta_k\right) c(t) + \sum_{k=1}^{K} \eta_k \gamma_k + \sum_{k=1}^{K} \eta_k \gamma_k \cos\left(f_k t + \omega_k^{(0)}\right).$$

The loglikelihood function is

$$\begin{aligned}
\ell(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{f}, \boldsymbol{\omega^{(0)}}; n) &= \sum_{i=1}^{n} \log\left(\theta(t_i)\right) - \int_0^T \theta(t)\, dt \\
&= \underbrace{\sum_{i=1}^{n} \log\left\{ \left(1 - \sum_{k=1}^{K} \eta_k\right) c(t_i) + \sum_{k=1}^{K} \eta_k \gamma_k + \sum_{k=1}^{K} \eta_k \gamma_k \cos\left(f_k t_i + \omega_k^{(0)}\right) \right\}}_{A} + \\
&\quad - \int_0^T \left[\left(1 - \sum_{k=1}^{K} \eta_k\right) c(t) + \sum_{k=1}^{K} \eta_k \gamma_k + \sum_{k=1}^{K} \eta_k \gamma_k \cos\left(f_k t_i + \omega_k^{(0)}\right) \right] dt \\
&= A - \left(1 - \sum_{k=1}^{K} \eta_k\right) \int_0^T c(t)\, dt - \sum_{k=1}^{K} \eta_k \gamma_k T + \\
&\quad - \sum_{k=1}^{K} \left[\eta_k \gamma_k \int_0^T \cos\left(2\pi(f_k t + \omega_k^{(0)})\right) dt \right]
\end{aligned}$$

$$
\begin{aligned}
&= A - \left(1 - \sum_{k=1}^{K} \eta_k\right) \sum_{j=1}^{N} c_j \Delta D_j - \sum_{k=1}^{K} \eta_k \gamma_k T + \\
&\quad - \sum_{k=1}^{K} \frac{\eta_k \gamma_k}{2\pi f_k} \left[\sin\left(2\pi(f_k T + \omega_k^{(0)})\right) - \sin\left(2\pi\omega_k^{(0)}\right)\right] \\
&= A - \left(1 - \sum_{k=1}^{K} \eta_k\right) \sum_{j=1}^{N} c_j \Delta D_j + \\
&\quad - \sum_{k=1}^{K} \eta_k \gamma_k \left[T + \frac{1}{2\pi f_k} \left[\sin\left(2\pi(f_k T + \omega_k^{(0)})\right) - \sin(2\pi\omega_k^{(0)})\right]\right] .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\ell(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{f}, \boldsymbol{\omega^{(0)}}; n) &= \sum_{i=1}^{n} \log\left\{\left(1 - \sum_{k=1}^{K} \eta_k\right) c(t_i) + \sum_{k=1}^{K} \eta_k \gamma_k + \sum_{k=1}^{K} \eta_k \gamma_k \cos\left(f_k t_i + \omega_k^{(0)}\right)\right\} + \\
&\quad - \left(1 - \sum_{k=1}^{K} \eta_k\right) \sum_{j=1}^{N} c_j \Delta D_j + \\
&\quad - \sum_{k=1}^{K} \eta_k \gamma_k \left[T + \frac{1}{2\pi f_k} \left[\sin\left(2\pi(f_k T + \omega_k^{(0)})\right) - \sin(2\pi\omega_k^{(0)})\right]\right] .
\end{aligned}
$$

Chapter 3

Data Analysis Using Multiscale Poisson Models

3.1 Introduction

This chapter applies the methodology of Chapter 2 on both simulated and real data. Through these data analyses we show that the multiscale methodology performs well in the analysis of neural spike trains. To estimate the variability of the multiscale estimators, quasi-likelihood bootstrap confidence intervals for the multiscale intensity function are developed. In an example, it is shown that the reconstruction quality of a complex intensity function demonstrates the ability of the multiscale methodology to crack the neural code. We also show, through an example, that the correlation coefficient among spike trains depends on the timescale.

3.2 Retinogeniculate synapse data analysis

We have employed the additive and multiplicative models of Chapter 2 on the data from a retinogeniculate synapse study. The lateral geniculate nucleus (LGN), which is located in the thalamus of the brain, receives input from retinal ganglion cells and is heavily involved in our perception of visual stimuli. The LGN also regulates the strength of the signals sent to the V1 area in visual cortex. Three types of cells exist in LGN: magnocellular (M cells), parvocellular (P cells), and koniocellular (K cells). Depending on their responses to the stimulus signal, each cell is either ON-center or OFF-center, (Nicholls et al., 2012, p.27) and Casagrande and Ichida (2011). It is known that only about half of the retinal spikes are transmitted to the cortex by the lateral geniculate nucleus (LGN), however, the unknown rules of such selections moti-

vate studies on the spike trains of the retinogeniculate synapse, Carandini et al. (2007); Sincich et al. (2007).

The data includes recordings from connected LGN and retinal ganglion cells (RGC) of a rhesus monkey. The subject was anesthetized and a neuromuscular blockade was established to prevent eye movement during recordings. Installing contact lenses, the eyes of the experimental subject were focused on a translucent tangent screen located 144cm away. The stimulus signal was a beam of diffused LED directed to the back of this screen at the centre of the receptive field. This signal was also restricted to the receptive field to avoid stimulating other retinal cells whose receptive fields may have overlapped with the cells being recorded. The light intensity changed constantly with a temporal power spectrum between 0.2 and 80Hz, producing approximately the luminance intensities measured when viewing natural scenes. The data includes repeated trials (5 seconds each) with the same stimulus signal across trials. For more details on the experimental methods and the recording process, refer to Sincich et al. (2007). The recorded LGN cells are either magnocellular or parvocellular. Combining the cell type with response type (ON-center of OFF-center), there are four possible combinations, from each of which we had one pair of physically connected RGC-LGN cells with repeated trials. Figure 3.1 shows the raster plot of the first 25 replications from each combination. The curve superimposed on the raster plot of the LGN replications represents the stimulus intensity. It is clear in Figure 3.1 that paired LGN-RGC cells have similar responses to a stimulus signal, with the RGC cell having a rate of almost twice as much as the LGN's [see Sincich et al. (2007)].

For clarity, notice that ON-centre parvocellular recordings will be fully analyzed here. Later in this chapter, we represent the multiscale fits and frequency estimates (Table 3.3) based on the recordings from the ON-centre parvocellular cell. Similar results to these ones for other pairs of neurons (OFF-centre and/or magnocellular), are provided in Section 3.4. We employ the recursive partitioning (RP) method, which sacrifices some computational efficiency compared to the RDP method. However, RP searches over a much richer class of recursive partitions, which turns out to be crucial in this case. The justification of this choice, is related to the *windowing effect*, which was discussed in Chapter 2. For computational convenience we assume that $N = 2^J,\ J = 1, 2, 3, \ldots$. Furthermore, the one-trial-at-a-time cross-validation process for the tuning parameters J and λ is based on a random sample of 35 trials from the 129 available replications. The search for optimum λ is over 100 equally spaced values on the closed interval $[0, 1.5]$, while the search for optimum N is over the set $\{2^1, 2^2, ..., 2^{10}\}$. Based on this scheme, we have used the formula introduced in Equation (2.4.5) on recordings from

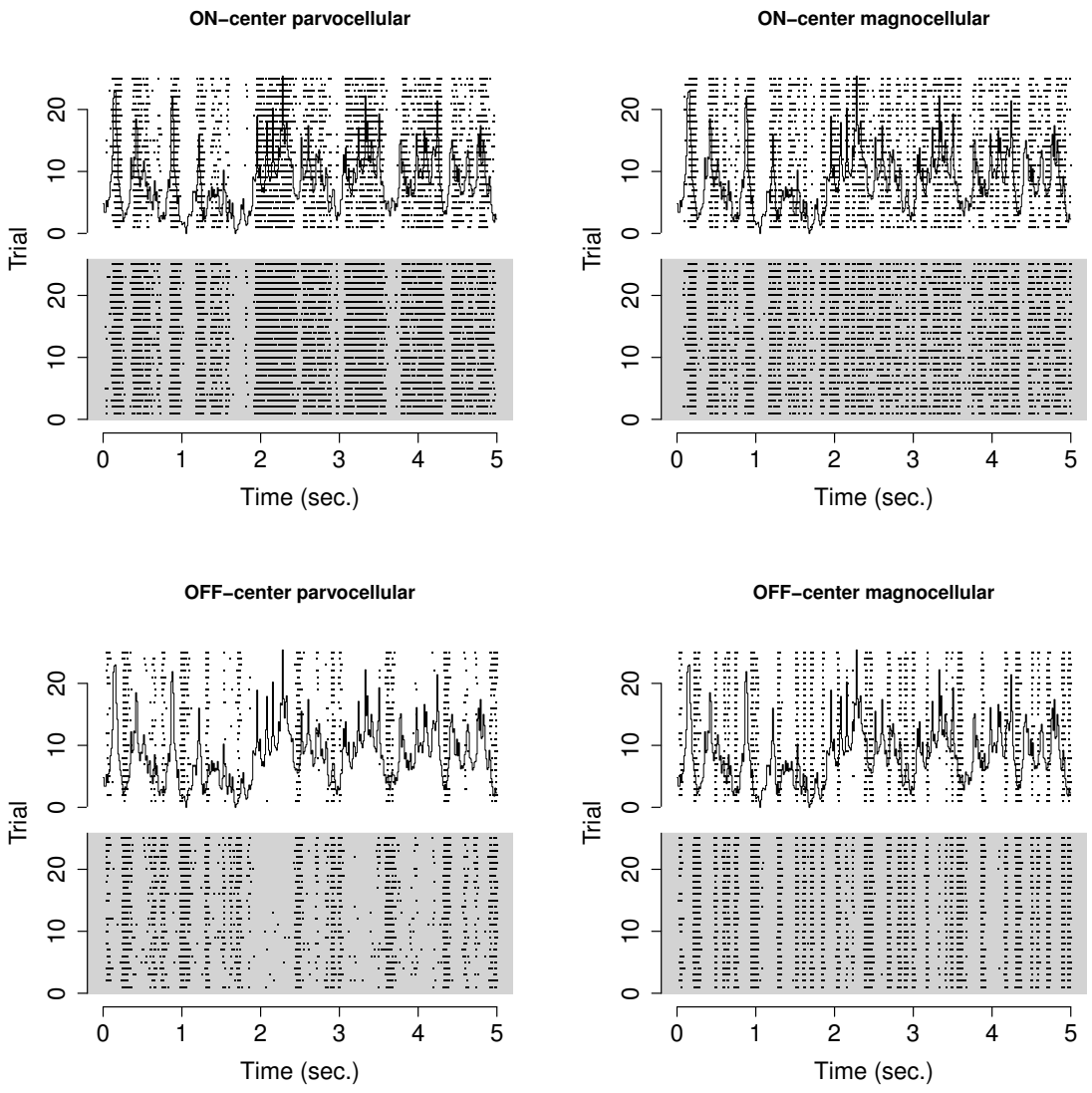


Figure 3.1: *The raster plot of the first 25 replications of each pair. The lower part of each panel (gray background) shows the RGC data while the upper panel plots the LGN data. The superimposed curve is the scaled stimulus intensity, which is the same across the trials for RGC and LGN neurons.*

repeated trials of paired LGN and RGC neurons and the optimal parameter pairs are presented in Table 3.1.

One interesting observation is that the estimated tuning parameters based on the retinal data (RGC) are quite close to those of its connected LGN neuron. Notice that

Table 3.1: *The estimated pair* (N^*, λ^*) *from Equation* (2.4.5) *for the two connected RGC and LGN cells.*

Cell	(N^*, λ^*)
RGC	$(2^6\ ,\ 0.030)$
LGN	$(2^6\ ,\ 0.015)$

$N^* = 2^6$ based on both the LGN and the RGC data. The λ^* values are also quite close to each other (given the range of values over which the optimization was done). This is consistent with the similarity in the spiking activity of these paired neurons plotted in the two panels of Figure 3.1. We have also noticed that the correlation coefficient between spike counts based on LGN and RGC data is negatively correlated with N, which shows that spike count correlation (a.k.a. "noise correlation") is also a multiscale phenomenon. This phenomenon has also been reported by Kass and Ventura (2006). Table 3.2 shows the average correlations for different values of N. Notice that although spike count correlation had no role in cross validation, $N^* = 2^6$ still results in a very high correlation between the spike counts of the two neurons. For a relatively recent study on the influential (experimental and physiological) factors on correlation measurements refer to Cohen and Kohn (2011).

Figure 3.2 summarizes the multiscale fits of the intensity function for the 129 trials of both RGC and LGN cells by plotting its average along with the 2.5% and 97.5% quantiles based on the repeated trials.

As mentioned at the end of Section 2.4.4 in Chapter 2, we can also perform the cross-validation based on the maximum similarity of the fit to the stimulus signal to explain most of the variability in the stimulus signal. The raster plots in the upper panels of Figure 3.1, particularly for the first few peaks of the stimulus signal, show some stimulus onset or latency, which is the time difference between the stimulus onset and the first spike. Spike latency Ventura (2004); Pawlas et al. (2010); Uzuntarla et al. (2012) is not a fixed number across trials and thus can be considered a random variable. This biological property of neurons has an important role in this particular setting for cross-validation. We have noticed that if the cross validation is applied on any given trial while adjusting for this latency, the estimate of the intensity function will capture a lot of the structure in the stimulus signal. Performing the cross validation across all trials with the average latency estimate produces a fit which smooths out a lot of the variability in the stimulus signal. Figure 3.3 plots the two scenarios.

Table 3.2: *Spike count correlation between RGC and LGN for different values of N. The correlations are the average of 129 trials with the standard deviations in brackets.*

N	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}
Correlation	1.00	0.99	0.99	0.98	0.97	0.95	0.93	0.89	0.82	0.70	0.59	0.48
(sd)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)

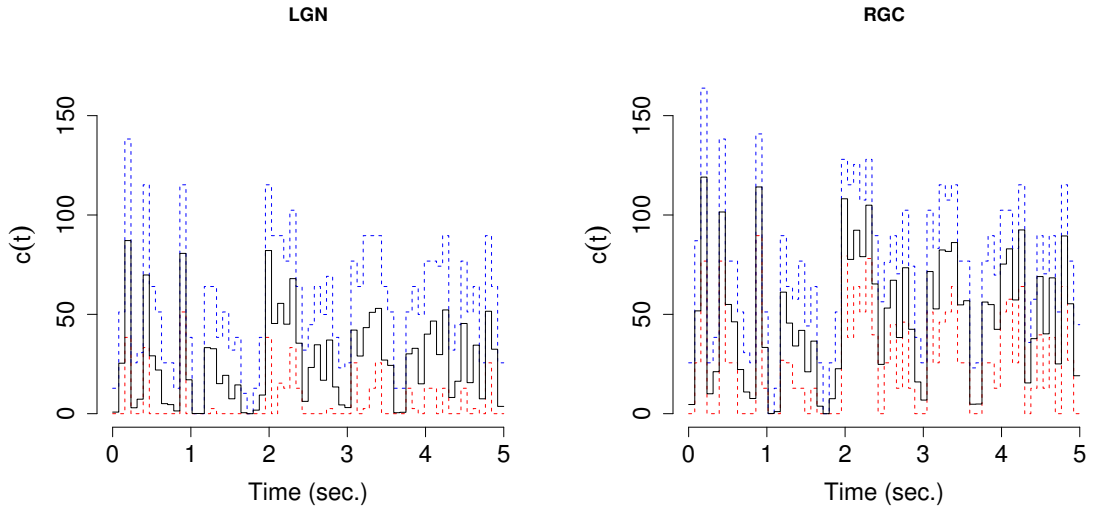


Figure 3.2: *The average of the multiscale fit of the intensity function across 129 trials. The blue and the red dashed lines represent, respectively, 97.5% and 2.5% quantiles based on the 129 repeated trials.*

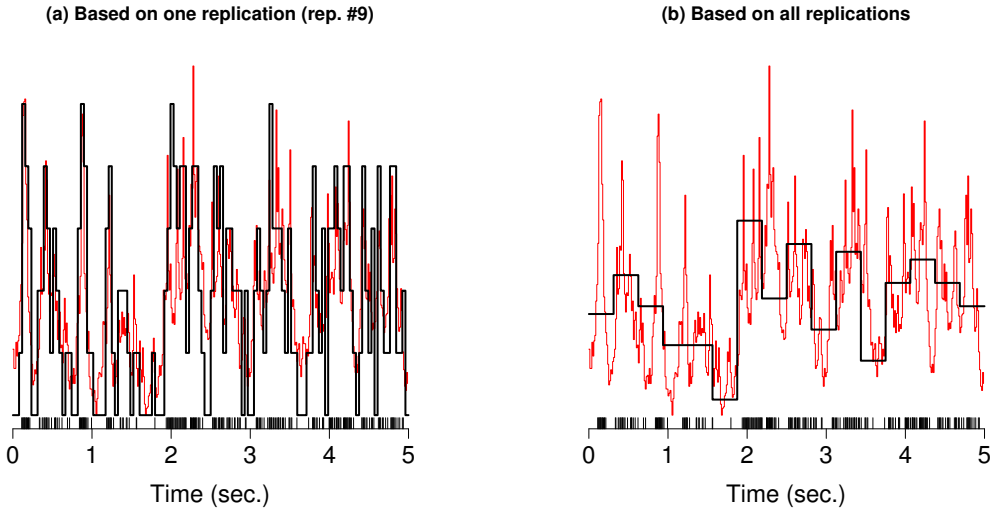


Figure 3.3: *The scaled fits based on the estimated tuning parameters. The fit in panel (a) is based on one replication while that of panel (b) is based on all of the 129 replications. The $\widehat{c}(t)$ function plotted in panel (b) is for the same data as in panel (a). The raster plots for the two panels are from the same replication. The stimulus intensity is shown in red.*

Figure 3.4 shows the multiscale fit associated with the pair $(N^*, \lambda^*) = (2^6, 0.01)$ for one randomly chosen replication (rep. #26), where the gray area represents a 95% point-wise confidence band based on the quasi-likelihood with variance function $V(\mu) = \phi_1 + \phi_2\mu$. In the variance function, μ is the mean, $V(\mu)$ is the variance as a function of the mean, and the two coefficients ϕ_1 and ϕ_2 determine the exact form of the variance function $V(\mu)$, and are estimated from the data through standard regression methods. We have noticed that the choice of the variance function greatly depends on the resolution parameter N (or equivalently J). For example if the "optimal" value for J was 7 as opposed to 6, then $V(\mu) = \phi_1\mu^{\phi_2}$ would have been a better choice for the variance function. This shows the multiscale nature of spike count variability, which has been previously addressed in Nelson (2002). To compute the confidence band we have simulated 500 spike trains from the estimated intensity function, which has been plotted in black in Figure 3.4, and used the 2.5% and the 97.5% sample quantiles. This band was then scaled according to the aforementioned variance function.

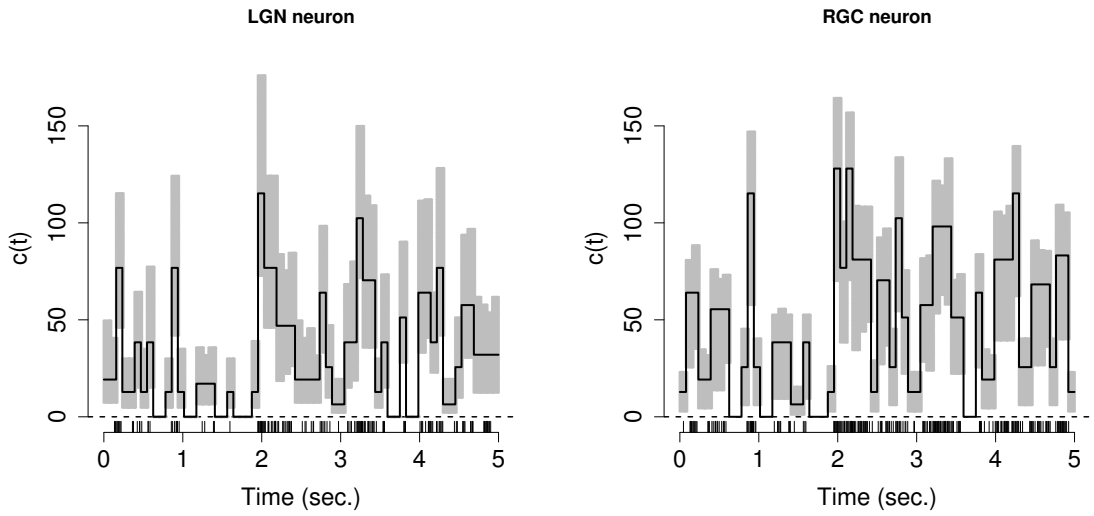


Figure 3.4: *The multiscale fit based on a randomly chosen replication (rep. #26). The gray area shows a* 95% *point-wise confidence band.*

Table 3.3 represents the 5 most common frequencies across the 129 trials of both RGC and LGN cells. We choose the frequencies which appear in at least half of the trials. Based on this table, we choose $K = 3$ periodic terms, as three frequencies appear in at least about 50% of trials.

The dominant frequency in the data (2.8Hz), lies in the δ-rhythm range. Although this

Table 3.3: *The 5 most repeated frequencies (in Hz) across the 129 trials. In brackets is the percentage of trials which reported the frequency.*

RGC	2.8(100%)	3.8(65.9%)	3.0(55%)	4.2(48.8 %)	5.6(46.5%)
LGN	2.8(98.5%)	3.8(55.8%)	5.6(48.8%)	4.2(31.0%)	5.4(17.1%)

is consistent with the result of Nuñez et al. (1992), where the presence of a δ-rhythm in the LGN of an anesthetized cat has been reported, notice that this could be due to the large peak at 2.81Hz in the temporal power spectrum of the stimulus intensity. Figure 3.5 plots the Fourier transform of the stimulus signal.

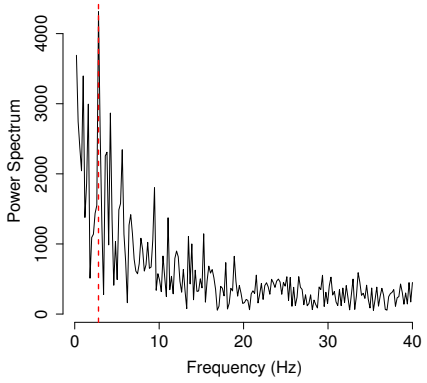


Figure 3.5: *The Fourier transform of the stimulus signal (the curve super imposed on the LGN trials in Figure 3.1). The red dashed line shows the maxima which is at 2.814Hz.*

Figure 3.6 shows the fit of the multiplicative, Equation (2.5.8), and the additive, Equation (2.5.9), models for a retinal ganglion cell (RGC) as well as its connected LGN neuron. The estimated parameters are presented in Table 3.4. Notice that the models are different in terms of complexity as $\eta_k \approx 0$ implies that the k^{th} periodic component can be omitted from the model. Since the models with fewer than three frequencies were all outperformed by the one with 3 periodic components, we only present the results of the models with $K = 3$ periodic components. We have also estimated the intensity function using Bayesian Adaptive Regression Splines (BARS) introduced in Dimatteo et al. (2001), which is a Bayesian free-knot curve fitting technique. The BARS fits for the RGC and LGN are also plotted in Figure 3.6. It is clear in this figure that relative to BARS, more details of the dynamics of the spiking activity are picked up by the

multiscale, multiple frequency models. Looking at panel (a) in Figure 3.3, it is clear that these extra details are related to the changes in the stimulus signal, which are smoothed out in the BARS fit.

Table 3.4: *Parameter estimates and 95% confidence intervals for the multiplicative and the additive models.*

Neuron	**Parameter**	**Multiplicative model**	**Additive model**
RGC	$\hat{\boldsymbol{f}}$	$(2.8, 3.8, 3.0)$	$(2.8, 3.8, 3.0)$
	$\hat{\boldsymbol{\omega}}^{(\mathbf{0})}$	$(0.454, 0.468, 0.400)$	$(0.454, 0.468, 0.400)$
	$\hat{\eta}_1$	$0.007 \quad (0.001, 0.02)$	$< 10^{-6}$
	$\hat{\eta}_2$	$< 10^{-6}$	$< 10^{-7}$
	$\hat{\eta}_3$	$< 10^{-8}$	$< 10^{-5}$
	$\hat{\gamma}_1$	$16.635 \quad (1.801, 32.909)$	17.872
	$\hat{\gamma}_2$	10.882	15.650
	$\hat{\gamma}_3$	11.831	11.749
	AIC	-1464.774	-1467.874
	AICc	-1422.774	-1434.941
	BIC	-1347.522	-1362.061
LGN	$\hat{\boldsymbol{f}}$	$(2.8, 3.8, 5.6)$	$(2.8, 3.8, 5.6)$
	$\hat{\boldsymbol{\omega}}^{(\mathbf{0})}$	$(0.417, 0.458, -0.106)$	$(0.417, 0.458, -0.106)$
	$\hat{\eta}_1$	$0.247 \quad (0.109, 0.427)$	$< 10^{-7}$
	$\hat{\eta}_2$	$0.391 \quad (0.165, 0.629)$	$< 10^{-8}$
	$\hat{\eta}_3$	$0.362 \quad (0.179, 0.575)$	$< 10^{-8}$
	$\hat{\gamma}_1$	$1.452 \quad (0.927, 2.033)$	13.889
	$\hat{\gamma}_2$	$0.669 \quad (0.348, 1.022)$	12.356
	$\hat{\gamma}_3$	$1.001 \quad (0.646, 1.398)$	14.441
	AIC	-835.566	-798.142
	AICc	-766.683	-763.086
	BIC	-692.575	-689.469

To test the goodness-of-fit, Brown et al. (2002) developed a Kolmogorov-Smirnov (KS) test based on the time-rescaling theorem, which is later corrected for time-discretization effects in Haslinger et al. (2010). Comparing the multiscale models with BARS, not only does the multiscale model provide a better fit to the LGN data, it is also computationally much faster than BARS. In this dataset, estimation of the multiscale function $c(t)$ took about 0.015 seconds, while the BARS fit took 21.5 seconds for the

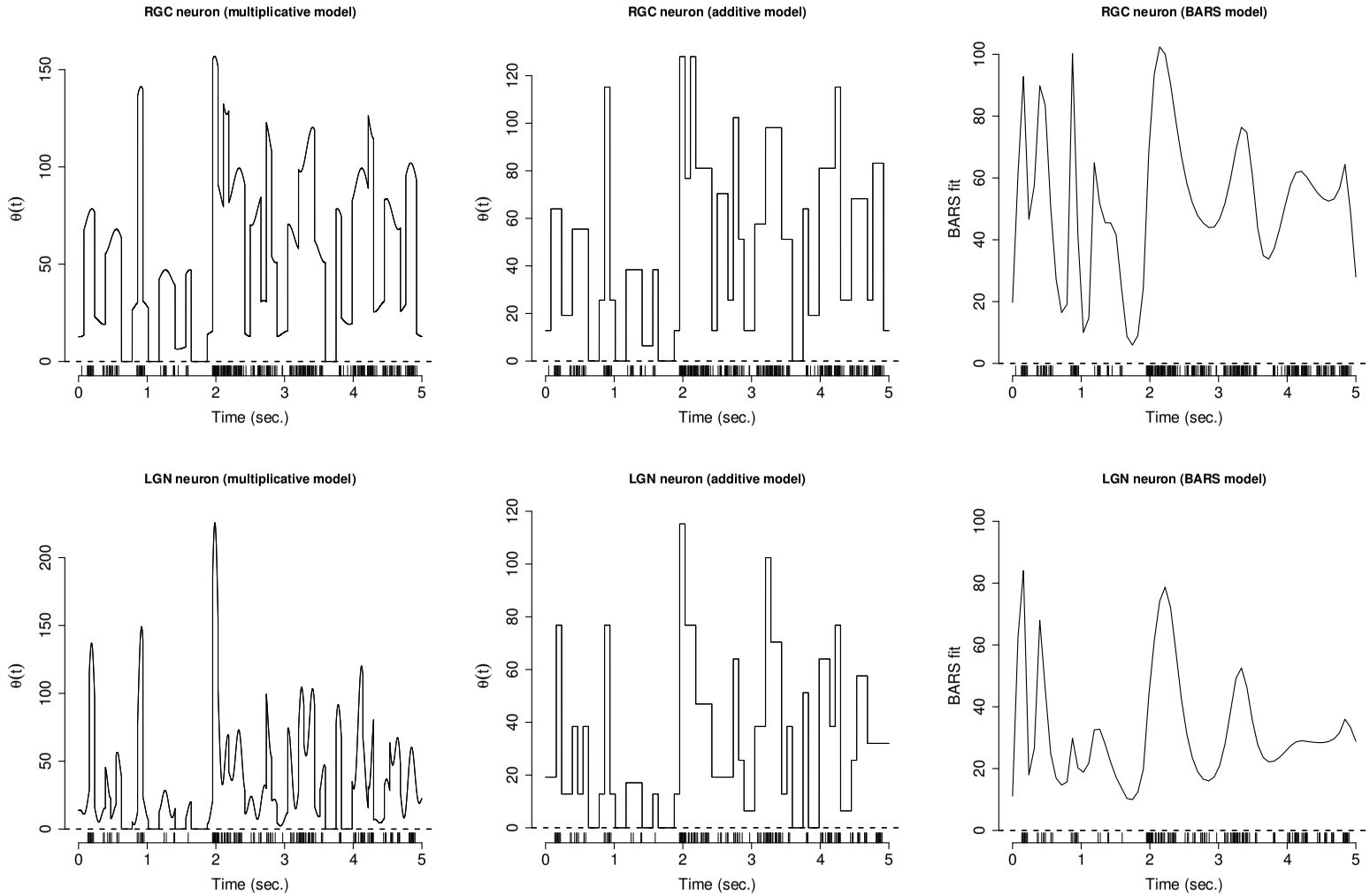


Figure 3.6: *The additive, multiplicative and BARS fits for the LGN and RGC data.*

RGC data, and 15.3 seconds for the LGN data on a 2.3GHz Intel Core 2 Duo with 2 GBs of memory. The optimization of the loglikelihood for $K = 3$ took another 5-6 seconds, which is still three to four times faster than the BARS method. The computational speed of the multiscale model is particularly important, because spike trains are usually very large datasets (hours of multiple trial recordings with ISIs in the order of milliseconds), and computational issues are quite common in this research area. Figure 3.7 shows the results of the goodness-of-fit test of Haslinger et al. (2010) on our models.

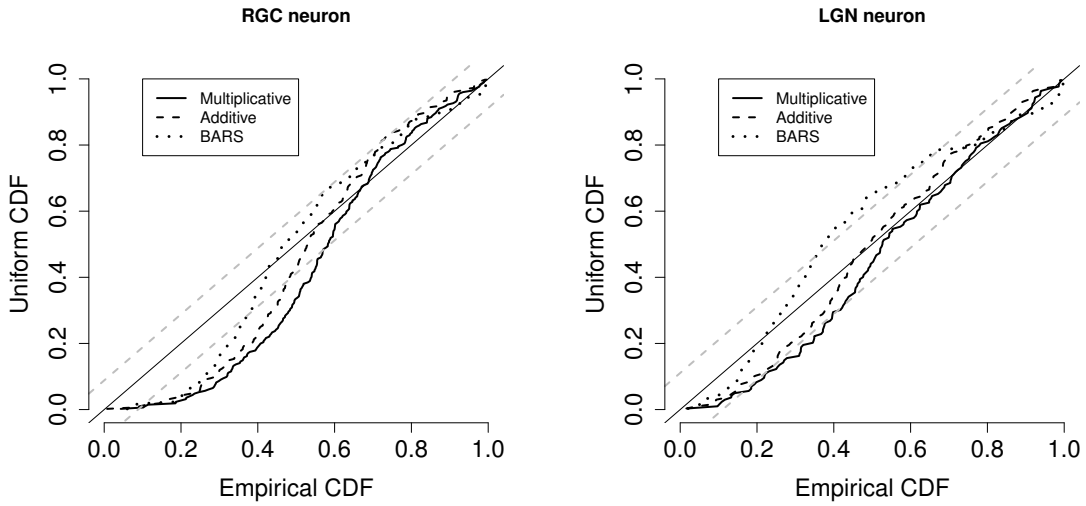


Figure 3.7: *Kolmogorov-Smirnov goodness-of-fit test for BARS, the additive, and the multiplicative models.*

In the additive models, none of the periodic terms contribute to the intensity function i.e., $\sum_k \eta_k \approx 0$ (notice that $\eta_k \geq 0$ for $k = 1, ..., K$). This is why the confidence intervals associated with the parameters of the additive model have not been computed. In fact, additive models suggest that $\widehat{\theta}(t) = \widehat{c}(t)$, the multiscale estimate with no periodic components. While model selection criteria (AIC, AICc and BIC) all suggest the multiplicative model for LGN, they all choose the additive model for the RGC data. Caution should be used in employing these models as none of them provide a particularly good fit for the RGC data; however, both multiplicative and additive models provide reasonable fits to the LGN data, both of which outperform the BARS fit (see Figure 3.7).

3.3 Discussion

In this chapter a dataset on the retinogeniculate synapse data (RGC and LGN neurons) was modeled using the multiscale methodology of Chapter 2 within the framework of an inhomogeneous Poisson process.

Based on the results presented in Table 3.1, it is clear that the time-scale of the spiking activity of the two retinal ganglion cell (RGC) and its connected LGN neuron are the same. We have also observed that the correlation coefficient between the spike counts of these two neurons is a multiscale quantity i.e., it depends on the bin size during which the spike are counted (see Table 3.2).

Pointwise bootstrap confidence bounds plotted in Figure 3.4 are the sample quantiles of 500 simulations of an inhomogeneous Poisson process with the intensity function plotted in black. These quantiles have been adjusted for the variance function of the associated quasi-likelihood. An alternative algorithm for computing the confidence band which we have found similar in terms of the final result is provided below.

Given the total number of spikes, n, and the number of intervals at the bottom of the recursive tree, N, the random vector of the spike counts at these N intervals follow a Multinomial distribution $\text{MN}(n, p_1, ..., p_N)$. The parameters $p_1, ..., p_N$ can be estimated from the initial spike train. Now, executing the following 5 steps, one can generate realizations of the intensity function whose sample quantiles can be used towards pointwise confidence band.

1. Generate $n^* \sim \text{Pois}(n)$
2. Generate $(n_1, ..., n_N) \sim \text{MN}(n^*, \widehat{p}_1, ..., \widehat{p}_N)$
3. Generate n_i samples from $\text{Uniform}(I_i)$, where I_i, $i = 1, 2, ..., N$ are the N subintervals at the bottom of the partitioning tree.
4. Aggregate all N vectors from step 3 together and sort the numbers in an increasing fashion.
5. Estimate the multiscale intensity function based on the simulated process from the previous step and go to step 1.

Notice that this algorithm assumes that the Poisson distribution is a good approximation for the distribution of the total number of spikes of a given trial. If this is not

a valid assumption, the variance of the estimation can be modified through a similar quasi-likelihood method employed earlier in this chapter.

We showed that in some cases, the proposed multiscale model for the intensity function of neural spike trains fitts the data better than the BARS model. Beside being computationally efficient, the multiscale model also outperforms BARS in terms of reconstruction of the stimulus signal.

3.4 Supplementary plots and frequency values for other neurons

3.4.1 Summary of the multiscale fits for other neurons

The following figures provide the average of multiscale fit for the other 3 paired neurons, whose data is plotted in Figure 3.1. These are OFF-centre M and P cells as well as an ON-centre P cell. Each combination includes multiple trials from an LGN neuron as well as its connected retinal ganglion cell. The coloured dashed lines show 2.5% and 97.5% quantiles based on the repeated trials within each time bin. Figures 3.8 and 3.9 plot the results.

We have extracted the five most common frequencies for all pairs of neurons in Tables 3.5-3.8. Notice that since both the size of cell (P-type vs M-type) and the response type of the cell (ON-centre vs OFF-centre) changes, the common frequencies are not completely the same, however, the majority of them are repeated in all cells. The RGC data is extracted from the input to the corresponding LGN neuron. In each table, RGC means the retinal ganglion cell connected to the LGN cell mentioned in that table.

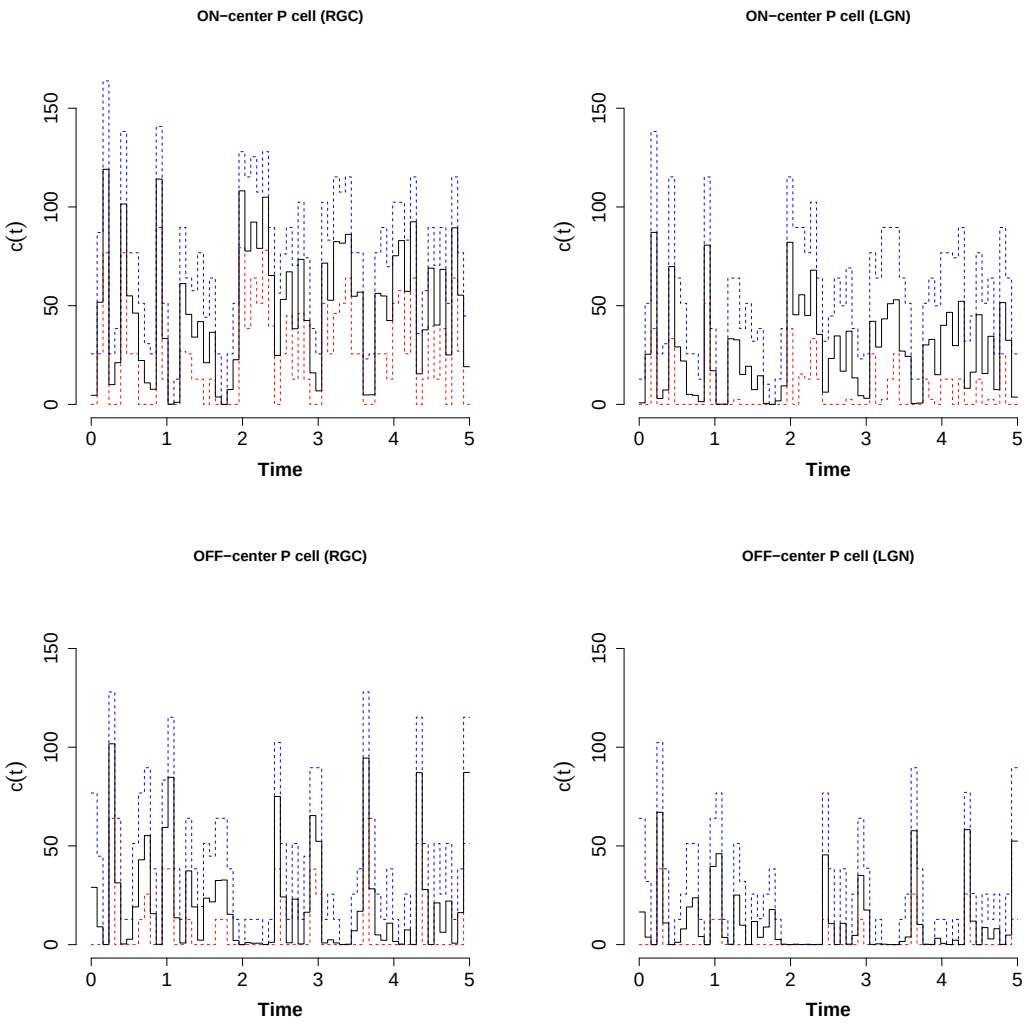


Figure 3.8: *The multiscale fit for Parvocellular cells. The red and blue dashed lines are the 2.5% and 97.5% quantiles based on the repeated trials.*

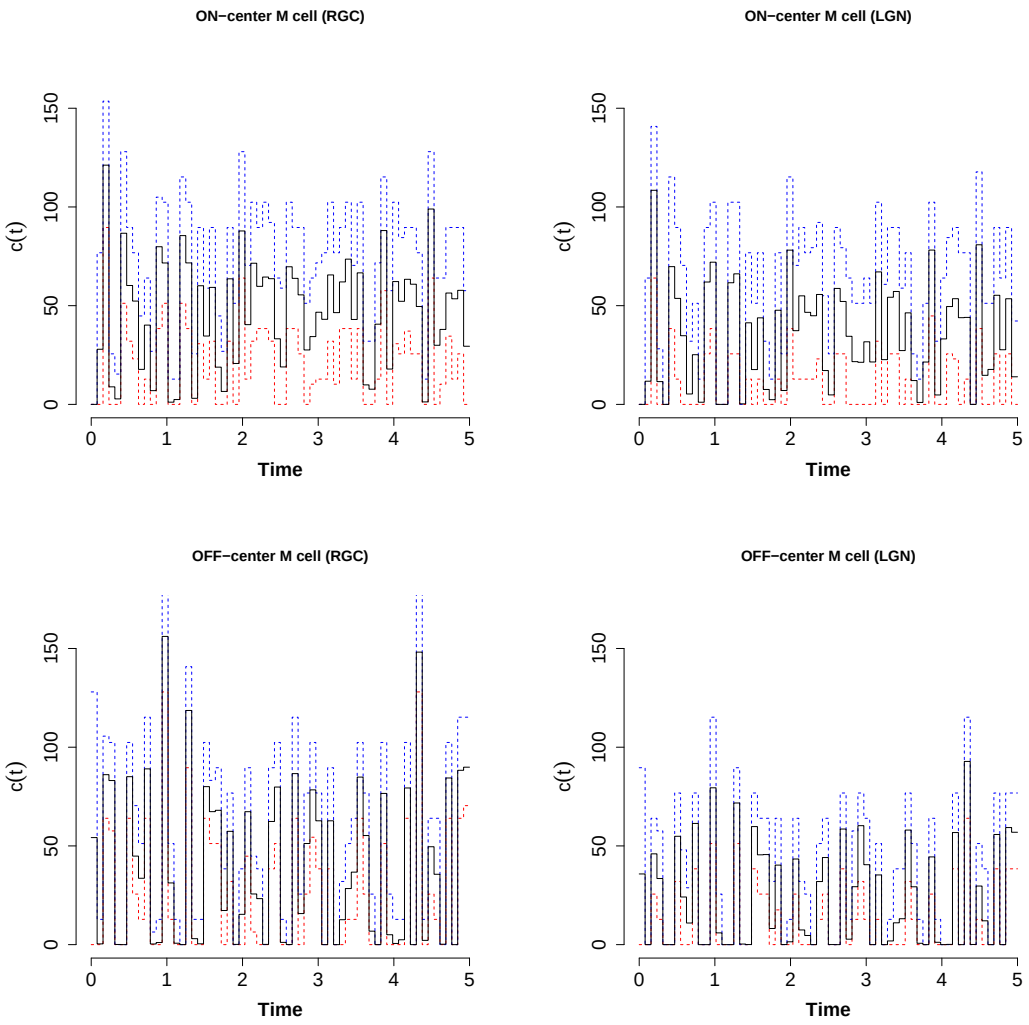


Figure 3.9: *The multiscale fit for Magnocellular cells. The red and blue dashed lines are the 2.5% and 97.5% quantiles based on the repeated trials.*

Table 3.5: *The 5 most repeated frequencies in Hz (and their percentages) across the 129 trials.*

RGC (ON-center P cell)	2.8(100%)	3.8(65.9%)	3.0(55%)	4.2(48.8 %)	5.6(46.5%)
LGN (ON-center P cell)	2.8(98.4%)	3.8(55.8%)	5.6(48.8%)	4.2(31%)	5.4(17.1%)

Table 3.6: *The 5 most repeated frequencies in Hz (and their percentages) across the 193 trials.*

RGC (ON-center M cell)	9.4(87.0%)	2.8(78.2%)	4.2(69.9%)	5.6(53.4%)	6.4(42.5%)
LGN (ON-center M cell)	2.8(83.9%)	9.4(69.9%)	4.2(63.2%)	5.6(54.9%)	6.4(47.7%)

Table 3.7: *The 5 most repeated frequencies in Hz (and their percentages) across the 400 trials.*

RGC (OFF-center P cell)	3(100%)	2.8(99.3%)	4.2(97.5%)	3.6(68.5%)	5.4(61.3%)
LGN (OFF-center P cell)	4.2(94.5%)	3(93.8%)	2.8(88.8%)	5.4(66.8%)	3.6(39.3%)

Table 3.8: *The 5 most repeated frequencies in Hz (and their percentages) across the 301 trials.*

RGC (OFF-center M cell)	11.6(97.7%)	2.8(56.5%)	3.6(55.5%)	5.6(49.2%)	4.2(47.5%)
LGN (OFF-center M cell)	11.6(98.7%)	15.4(76.7%)	12.2(65.4%)	7.0(39.2%)	9.2(35.5%)

Chapter 4

Univariate Skellam Process With Resetting

4.1 Introduction

This chapter introduces the Skellam process and studies its properties in both the homogeneous and inhomogeneous cases. In order to study neural spike trains in the framework of a Skellam process, we also introduce the *Skellam Process with Resetting (SPR)*. Introduction of this process (SPR) and its application in the analysis of neural spike trains is one of the major contributions of this dissertation. The Skellam process, which is the difference between two independent Poisson processes, codes information with two parameters/time-varying functions and is biologically justifiable. In fact, motivated by the process of neural integration, (Carlson, 2007, 62-63), the spikes in the observation interval $(0, T]$ are modelled as the records of the so-called Skellam process. The only difference between the records of a Skellam process and SPR is that in the latter process, we add a short period of "resetting" after each spike, which is motivated by the refractoriness of neurons and independence considerations. The two Poisson processes defining the Skellam process play the role of the integrated presynaptic *inhibitory* and *excitatory* effects. Motivated by the time-dependent behaviour of neural spiking activity, which was discussed in the previous chapter, we also study the properties of this Skellam model under different time-scales.

4.1.1 Neural inhibition

Modelling neural spiking activity in an inhomogeneous Poisson process framework is a common approach in the literature. However, one of the weaknesses of Poisson process is its poor performance in approximating the interspike interval (ISI) distribution,

Kass and Ventura (2001); Brown et al. (2002, 2003). Furthermore, there are periods of neural inactivity, during which the only conclusion within Poisson process framework would be that the intensity function is zero (or close to zero). While a zero intensity is not an incorrect conclusion, it does not provide rich biological insight to the time intervals during which the neuron does not initiate spikes. Neural integration is the process by which a neuron aggregates the synaptic potentials. If the integrated postsynaptic potential reaches the threshold of excitation (≈-60mV) at time t_0, a spike is released (Carlson (2007) pp.45, 62-63). During the spiking inactivity periods the neuron is not necessarily resting, rather it is inhibited by the strong inhibitory postsynaptic potentials. This means that there are still many chemical activities at its synapses (high amount of input), but the integrated postsynaptic potential does not meet the threshold of excitation. The more inhibited, the longer the inactivity period, hence, there could be autocorrelation among interspike intervals (ISIs). The authors de Ruyter van Steveninck et al. (1997) showed that the spike trains which show exponential ISIs and seem to be well-approximated through Poisson process, are, in fact, more reproducible than the simulated Poisson spike trains. In other words, the individual trials in the raster plot of real data look more like each other compared to that of simulated data from Poisson process. The dependence among ISIs, along with the limitations in the correlation structure of Poisson models, Kocherlakota and Kocherlakota (1992), motivate the idea of utilizing a more detailed model for studying neural spike trains.

4.2 Skellam distribution and Skellam process

Consider two independent random variables $X^{(1)}$ and $X^{(2)}$ where $X^{(i)} \overset{\text{ind.}}{\sim} \text{Pois}(\lambda_i)$. $X = X^{(1)} - X^{(2)}$ is called a *Skellam random variable* with parameters λ_1 and λ_2, denoted by $X \sim Sk(\lambda_1, \lambda_2)$. Studied by Skellam (1946), the probability mass function of X is

$$\begin{aligned} p_{\lambda_1,\lambda_2}(x) &= \exp\{-(\lambda_1+\lambda_2)\}\sum_{y=0}^{\infty}\frac{\lambda_1^{x+y}\lambda_2^{y}}{(x+y)!\ y!} \\ &= \left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right)^{x}\exp\{-(\lambda_1+\lambda_2)\}B_x(2\sqrt{\lambda_1\lambda_2}), \end{aligned}$$

where $x = 0, \pm1, \pm2, ...$ and $B_x(\cdot)$ is the modified Bessel function of the first kind (Abramowitz and Stegun, 1972, p.375),

$$B_x(t) = \sum_{n=0}^{\infty}\frac{(t/2)^{x+2n}}{n!(x+n)!}\ . \tag{4.2.1}$$

The probability generating function of a Skellam distribution with parameters λ_1 and λ_2 is

$$\begin{aligned} G_X(w) &= \mathrm{E}\left(w^X\right) \\ &= \exp\Big[-(\lambda_1+\lambda_2)+\lambda_1 w+\frac{\lambda_2}{w}\Big] \qquad \text{for } w \neq 0. \end{aligned} \tag{4.2.2}$$

It is also easy to check that if $X_i \sim Sk(\lambda_{1i}, \lambda_{2i})$, $i = 1, ..., n$ are independent Skellam random variables, then

$$X_1 \pm X_2 \pm \cdots \pm X_n \sim Sk\Big(\lambda_{11} \pm \lambda_{12} \cdots \pm \lambda_{1n}\ ,\ \lambda_{21} \pm \lambda_{22} \cdots \pm \lambda_{2n}\Big).$$

The special case of the Skellam distribution where $\lambda_1 = \lambda_2$ was first studied by Irwin (1937). Based on the cumulants of Poisson random variables, Skellam (1946) addressed the distribution of the difference between two independent Poisson random variables, which is now called *Skellam distribution*. Calling it the Poisson difference distribution, Alzaid and Omair (2010) provided the maximum likelihood estimates of the parameters λ_1 and λ_2 based on the properties of the Bessel function and studied their asymptotic distribution and compared them to the moment estimates. Karlis (2003) used a bivariate Poisson model for analysis of sports data and used Skellam distribution as the distribution of the difference between the scores of two teams. Poppe et al. (2008) used the Skellam distribution for the illumination allowance in a video surveillance study. For some applications of Skellam distribution/process in finance as well as the definition and some properties of the Skellam-Lèvy process see Barndorff-Nielsen et al. (2012). For some neuroscience applications of Skellam distribution see Shin et al. (2010), where they modelled changes in the firing activity of the neurons before and after a task through a Skellam distribution.

4.2.1 Homogeneous Skellam process

Modelling discrete-valued price changes, the Skellam process was first introduced in Barndorff-Nielsen et al. (2010) as the Skellam-Lèvy process. Studying the cumulants of the "standard Skellam process", in which $\lambda_1 = \lambda_2 = 0.5$, Barndorff-Nielsen et al. (2010, 2012) call this process a "discrete-valued analogy of Brownian motion." We define the Skellam process in the classic framework of Poisson process as follows;

Definition 4.1 *An integer-valued process* $\{S(t) : t \geq 0\}$ *is said to be a* homogeneous Skellam process *with nonnegative parameters* (λ_1, λ_2) *if*

1. *S(0)=0,*

2. *For all values* $t_1 < t_2$*, the random variable* $S(t_2) - S(t_1)$ *is independent of the times of incidents during* $[0, t_1]$ *(independent increments),*

3. $Pr\Big[S(t+h) - S(t) = 0\Big] = 1 - (\lambda_1 + \lambda_2)h + o(h)$,

4. $Pr\Big[S(t+h) - S(t) = 1\Big] = \lambda_1 h + o(h)$,

5. $Pr\Big[S(t+h) - S(t) = -1\Big] = \lambda_2 h + o(h)$,

6. $Pr\Big[\big|S(t+h) - S(t)\big| \geq 2\Big] = o(h)$,

as $h \to 0$.

Alternatively, Skellam process can be defined as follows,

Definition 4.2 *An integer-valued process* $\{S(t) : t \geq 0\}$ *is said to be a* homogeneous Skellam process *with nonnegative parameters* (λ_1, λ_2) *if*

1. *S(0)=0,*

2. $\{S(t), t \geq 0\}$ *has independent increments,*

3. *For all values* $t, s \geq 0$, $S(t+s) - S(s) \sim Sk(\lambda_1 t, \lambda_2 t)$.

Looking at Definition 4.2, the distribution of the increments is Skellam, which is an infinitely divisible distribution. More specifically, it is a Lèvy process. This guarantees the existence of a stochastic process which satisfies the conditions of the definition.

To the best of our knowledge, defining the Skellam process in this form was first introduced in neuroscience contexts in Ramezan et al. (2010) and was employed later to model real data (from visual cortex) in Ramezan et al. (2012). The results of these research studies are presented in this chapter and the next one.

Theorem 4.1 *Definitions 4.1 and 4.2 are equivalent.*

Proof: The proof is provided in Section 4.8.

Theorem 4.2 *Let* $N_i = \{N_i(t),\ t \geq 0\},\ i = 1, 2$ *be two independent Poisson processes with rates* λ_1 *and* λ_2*, respectively. The stochastic process* $\{S(t) = N_1(t) - N_2(t), t \geq 0\}$ *is a Skellam process with nonnegative parameters* (λ_1, λ_2).

Proof: The proof is provided in Section 4.8.

Figure 4.1 displays a realization of two Skellam processes derived from independent Poisson processes $\{N_i(t),\ t \geq 0\}$, $i = 1, 2$.

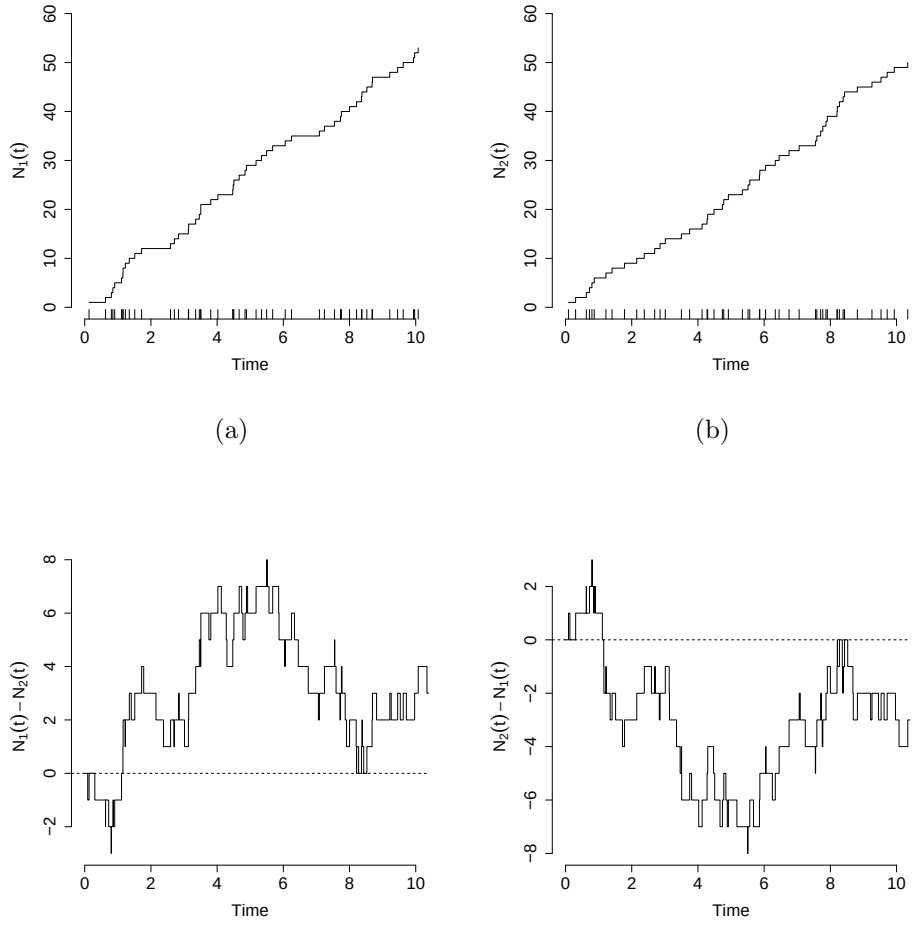


Figure 4.1: *Examples of homogeneous Skellam process. (a)&(b): Homogeneous Poisson processes N_1 and N_2 with equal parameters $\lambda = 5$ per unit time - (c)&(d): Skellam processes based on N_1 and N_2. (c): $\{S(t) = N_1(t) - N_2(t), t \geq 0\}$ and (d):$\{S(t) = N_2(t) - N_1(t), t \geq 0\}$. The rug plots in (a) and (b) show the event times.*

4.2.2 Inhomogeneous Skellam process

The homogeneous Skellam process defined above is a stationary process. We define the inhomogeneous Skellam process, or alternatively Skellam process with time-varying parameters, as follows.

Definition 4.3 *An integer-valued process $\{S(t) : t \geq 0\}$ is said to be an* inhomogeneous Skellam process *with time-varying parameters $\lambda_1(t)$ and $\lambda_2(t)$, $t \geq 0$ if*

1. *S(0)=0,*

2. $\{S(t), t \geq 0\}$ *has independent increments,*

3. $Pr\Big[S(t+h) - S(t) = 0\Big] = 1 - \Big(\lambda_1(t) + \lambda_2(t)\Big)h + o(h),$

4. $Pr\Big[S(t+h) - S(t) = 1\Big] = \lambda_1(t)h + o(h),$

5. $Pr\Big[S(t+h) - S(t) = -1\Big] = \lambda_2(t)h + o(h),$

6. $Pr\Big[\big|S(t+h) - S(t)\big| \geq 2\Big] = o(h).$

as $h \to 0$.

Alternatively, it can be defined as,

Definition 4.4 *An integer-valued process* $\{S(t) : t \geq 0\}$ *is said to be an* inhomogeneous Skellam process *with time-varying parameters* $\lambda_1(t)$ *and* $\lambda_2(t)$, $t \geq 0$ *if*

1. $S(0) = 0,$

2. $\{S(t), t \geq 0\}$ *has independent increments,*

3. *For all values* $t, s \geq 0$, $S(t+s) - S(s) \sim Sk\Big(\int_s^{t+s} \lambda_1(y)\, dy\ ,\ \int_s^{t+s} \lambda_2(y)\, dy\Big)$.

Similar to the homogeneous case, the existence of such process is guaranteed due to the infinite divisibility of the Skellam distribution for the increments. Of course, this is tied to the fact that Skellam process is, in fact, a Lèvy process.

Theorem 4.3 *Definitions 4.3 and 4.4 are equivalent.*

Proof: The proof is provided in Section 4.8.

The inhomogeneous version of Theorem 4.2 is the following,

Theorem 4.4 *Let* $N_i = \{N_i(t),\ t \geq 0\}$, $i = 1, 2$ *be two independent inhomogeneous Poisson processes with time-varying parameters* $\lambda_1(t)$ *and* $\lambda_2(t)$, $t \geq 0$, *respectively. The stochastic process* $\{S(t) = N_1(t) - N_2(t), t \geq 0\}$ *is an inhomogeneous Skellam process with time-varying parameters* $\lambda_1(t)$ *and* $\lambda_2(t)$.

Proof: The proof is similar to that of Theorem 4.2.

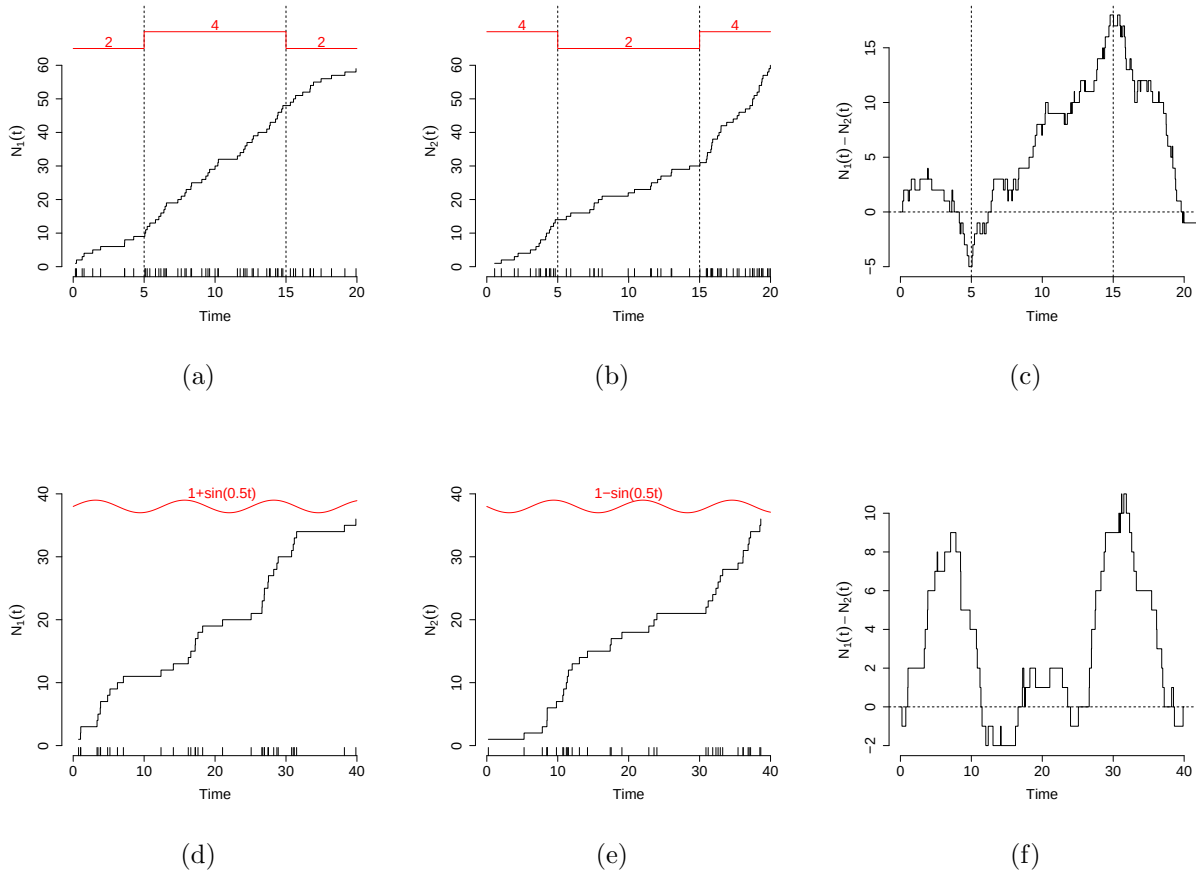


Figure 4.2: *Examples of inhomogeneous Skellam process. (a),(b),(d),(e): Inhomogeneous Poisson processes with their intensity functions in red. - (c)&(f): Inhomogeneous Skellam processes based on the pairs (a),(b) and (d),(e). (c): (a)-(b) and (f):(d)-(e)*

Figure 4.2 plots two realizations from inhomogeneous Skellam processes.

We will see that the spike times can be modelled as the record times[1] of a Skellam process, and in fact, the underlying Skellam process is not observable. In particular, the negative part of Skellam process, during which the neuron may be strongly inhibited, is not observable. It was mentioned in Chapter 1 that according to neural integration, a spike is initiated when the excitatory postsynaptic potentials "dominate" the inhibitory ones. Therefore, it seems logical to think of the records of the Skellam process as the spike times, as a record occurs every time the process $\{N_1(t), t \geq 0\}$ (aggregated excitatory input) "dominates" $\{N_2(t), t \geq 0\}$ (aggregated inhibitory input), i.e. when

[1]The record at time t is the maximum of the process up to and including time t.

$N_1 > N_2$. Therefore, the only observable part of the process are these record times, and the rest of the process is observed as 0. This matter, along with the effect of the refractory period, motivate introducing a new process which will be called *Skellam process with resetting*.

4.3 Skellam process with resetting (SPR)

We would like to study spike trains in the Skellam process framework. Analogous to the integrated spiking activity of excitatory and inhibitory presynaptic neurons are, respectively, $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$. Therefore, $\{S_0(t) = N_1(t) - N_2(t), t \geq 0\}$ plays the role of a jump process version of a neural integration process. Notice that $\{S_0(t), t \geq 0\}$ introduces negative values, which can (but do not necessarily have to) correspond to the resting or inhibition periods (depolarization), during which the integrated inhibitory postsynaptic potential dominates the integrated excitatory ones, i.e. $S_0(t) < 0$. However, as was mentioned before, these negative values are not observable.

Every time a neuron fires, most of its chemical mechanisms and the membrane potential reset during the refractory period (Nicholls et al., 2012, 120-121). Recall that the (absolute) refractory period is a short interval after a spike (1-3 milliseconds), during which the initiation of a second spike is not possible for the neuron. To incorporate this biological fact (resetting of the chemical mechanisms) into our model, after each record of the Skellam process (spike time) the process is reset, i.e., its value is brought back to zero.

This resetting, which is completely motivated by the neurophysiology of neurons, will bring mathematical convenience to our modelling. Notice that when the process resets, its path goes back to 0, giving it a fresh start. This motivates the modelling assumption that conditional on the n^{th} resetting time, i.e. the n^{th} spike time, the $(n+1)^{\text{th}}$ spike time is independent of the rest of the spiking history. In other words, the only history which is taken into consideration is the elapsed time from the previous spike. In Chapter 5, where the likelihood function based on spike trains is derived, we will see that this assumption of conditional independence simplifies the derivation of the likelihood function. This modelling assumption is, essentially, a similar idea to the Inhomogeneous Markov Interval (IMI) model introduced in Kass and Ventura (2001), where they formulated the conditional intensity, $\theta(t|H_t)$ in Equation (4.3.9), as a function of the experimental clock and the time elapsed since the previous spike. However, SPR and IMI have some fundamental differences. SPR is much more biologically justifiable than most of the common statistical approaches in modelling the conditional

intensity function of neural activity. The dynamics of sub-threshold neural activity (below the threshold of excitation) is captured through the difference $\lambda_1 - \lambda_2$, or in an inhomogeneous process framework through the difference $\lambda_1(t) - \lambda_2(t)$.

On the other hand, the fact that the spike times and the records of the Skellam process agree in this model is due to the fact that a record occurs every time $N_1(t)$ dominates $N_2(t)$, which is analogous to the domination of inhibitory potentials by the excitatory ones. It is important to emphasize that using records is motivated by the biological law of neural integration. Provided below is the formal definition of Skellam process with Resetting (SPR).

Definition 4.5 *Let R_t be the record of the Skellam process $S_0 = \{S_0(t), t \geq 0\}$ on the interval $[0, t)$ with $S_0(0) = 0$, i.e.,*

$$R_t = \max\Big\{S_0(t^*) : \ t^* < t\Big\}.$$

Skellam Process with Resetting (SPR) is defined to be

$$S = \Big\{S_0(t) - R_t, \ t \geq 0\Big\}.$$

For data analysis and estimation purposes, we will discretize time into very small intervals, which consequently, translates the "exact" spike times into a sequence of zeros and ones depending on whether or not a spike exists in each of these small intervals. Let $X^{(1)}$ and $X^{(2)}$ be independent Poisson processes with respective rates λ_1 and λ_2 observed over the interval $(0, T]$. Suppose this time interval is divided into k subintervals of length h, i.e.,

$$(0, T] = \bigcup_{i=1}^{k} \Big((i-1)h \, , \, ih\Big]. \tag{4.3.3}$$

Define $X_i^{(1)}$ and $X_i^{(2)}$ to be event counts in the i^{th} subinterval. Define X_i to be

$$X_i = \begin{cases} +1 & X_i^{(1)} > X_i^{(2)}, \\ 0 & X_i^{(1)} = X_i^{(2)}, \\ -1 & X_i^{(1)} < X_i^{(2)}. \end{cases} \tag{4.3.4}$$

Notice that $X_i \neq X_i^{(1)} - X_i^{(2)}$, but as h goes to zero (or k goes to infinity), X_i converges to $X_i^{(1)} - X_i^{(2)}$. For a given value of h limiting to 0, the probability function of X_i is

x_i	-1	0	1
$\Pr(X_i = x_i)$	p_-	p_0	p_+

where

$$p_{-} = \lambda_2 h + o(h), \tag{4.3.5}$$
$$p_{0} = 1 - (\lambda_1 + \lambda_2)\, h + o(h), \tag{4.3.6}$$
$$p_{+} = \lambda_1 h + o(h), \tag{4.3.7}$$

which are the same as the probabilities listed in Definition 4.3. Notice that right after each record, the Skellam process is forced to zero by Definition 4.5. Therefore, in the discretized version we have,

$$\Pr\Big(X_{i+1} = 0 \,\Big|\, \text{a record of } S_0 \text{ occurred in } \big((i-1)h, ih\big]\Big) = 1, \tag{4.3.8}$$

which imposes the resetting after each spike (record of S_0), and brings back the values of the Skellam process S_0 to 0.

A sample *path* or *trajectory* of the SPR $\{S(t), t \geq 0\}$ is shown in Figure 4.3. Starting at the origin, at each step the random variable increases (+1), decreases (-1) or stays at the same value (0). The black dots represent the spike times and the red lines show the forced-to-zero periods after each spike.

Assuming that $\{N_0(t), t \geq 0\}$ is the counting process associated with S_0, i.e. it counts the number of records of S_0, the conditional intensity function is

$$\theta(t|H_t) = \lim_{h \to 0} \frac{\Pr\big(N_0(t+h) - N_0(t) = 1 \big| H_t\big)}{h} \tag{4.3.9}$$

where $H_t = \{t_1, t_2, ..., t_{N(t)}\}$ is the history of the process up to time t. If the counting process N_0 is Poisson, then it will be independent of the history H_t. However, the resetting of the process in SPR brings dependence among consecutive spikes.

4.4 Skellam process and Markov chain

Discretizing time in Equation (4.3.3)'s fashion, the Skellam process with resetting is a Markov Chain with the state space $\{1, 0, -1, -2, ...\}$ and the Transition Probability Matrix (TPM)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ p_{+} & p_{0} & p_{-} & 0 & 0 & \dots \\ 0 & p_{+} & p_{0} & p_{-} & 0 & \dots \\ 0 & 0 & p_{+} & p_{0} & p_{-} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}. \tag{4.4.10}$$

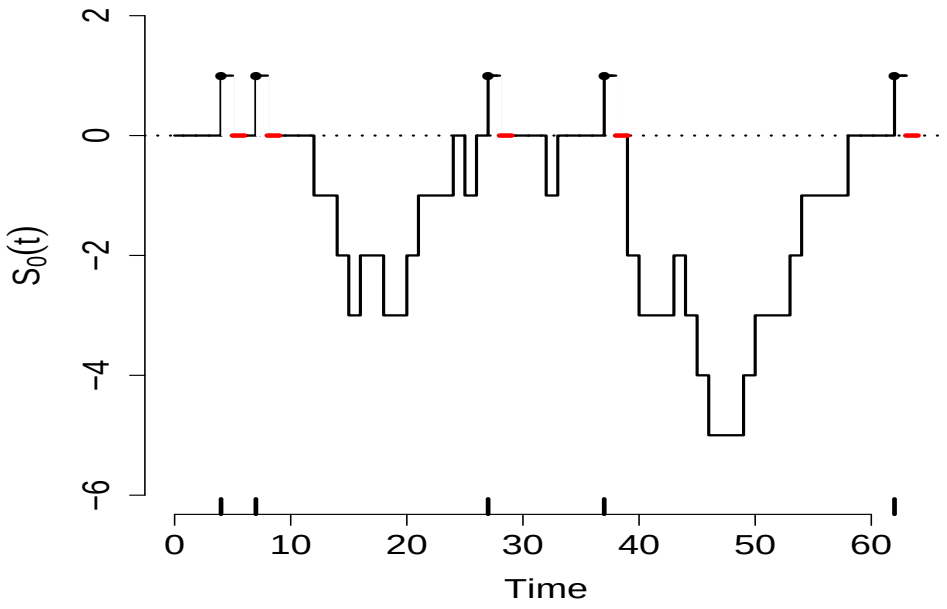


Figure 4.3: *A toy example plotting a spike train of 5 spikes and a possible trajectory of the process. The black dots represent spike times and the red short lines are associated with the resetting of the process.*

Since all states of this chain communicate, it is irreducible. Under the condition $\lambda_1 > \lambda_2$ the chain is positive recurrent. Furthermore, the diagonal elements $p_0 > 0$ imply the chain is aperiodic. Solving the system of equations

$$\pi_j = \sum_{i=1}^{\infty} \pi_i P_{ij} \quad j = 1, 2, ...$$

subject to $\sum_j \pi_j = 1$, the equilibrium distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ of this ergodic chain is

$$\begin{aligned} \pi_1 &= \frac{p_+ - p_-}{1 + p_+ - p_-}, \\ \pi_i &= \left(\frac{p_-}{p_+}\right)^{i-2} \frac{\pi_1}{p_+} \quad i = 2, 3, ... \end{aligned} \tag{4.4.11}$$

where the states $\{1, 0, -1, ...\}$ are mapped to $i \in \{1, 2, ...\}$ i.e., π_i is the long-run proportion of visits to the state $(2 - i)$. This limiting distribution is in the form of Power Series Distribution (PSD) which was introduced initially in Noack (1950) and

the properties of its cumulants and factorial cumulants were later studied by Khatri (1959). Patil (1962) generalized the results to the truncated version of PSD and called it Generalized Power Series Distribution (GPSD).

Two special cases of $\boldsymbol{\pi}$ are when p_- approaches either of the limiting values 0 or p_+. We have

$$\pi_1 \nearrow \frac{p_+}{1+p_+} \text{ as } p_- \searrow 0$$
$$\pi_2 \nearrow \frac{1}{1+p_+} \text{ as } p_- \searrow 0$$

This is consistent with the time-discretized version of the "Poisson process with resetting," where most of the visits are to the state 0 and occasional visits to state 1. Now let us study the case where $p_- \to p_+$. In this case, the condition for positive recurrence, $p_- < p_+$, fails at the limit and we have $\pi_i \to 0$, $i = 1, 2, ...$, which means that all states are transient and the equilibrium distribution does not exist.

It is noteworthy that based on the equilibrium (stationary) distribution $\boldsymbol{\pi}$, one crude estimate for the expected number of spikes in the observation period $(0, T]$ is $\pi_1(T/h)$. As $h \to 0$, this estimator converges to $(\lambda_1 - \lambda_2)T$, which is the mean value of a homogeneous Skellam process over $(0, T]$ with parameters λ_1 and λ_2.

4.4.1 Over dispersion

Let M_i be the random variable representing the value of the Skellam process with resetting at time i. Let us call this random variable under equilibrium M. Then the Moment Generating Function (MGF) of M is

$$\begin{aligned} E\left(e^{tM}\right) &= \sum_{i=1}^{\infty} \pi_i \, e^{(2-i)t} \\ &= \pi_1 \, e^t + \frac{\pi_1}{p_+} \sum_{i=2}^{\infty} e^{t(2-i)} \left(\frac{p_-}{p_+}\right)^{i-2} \\ &= \pi_1 \left(e^t + \frac{1}{p_+ - p_- \, e^{-t}}\right) \end{aligned} \tag{4.4.12}$$

Based on the MGF (4.4.12) we calculate the first two moments,

$$\begin{aligned} E(M) &= \frac{\partial}{\partial t} E\left(e^{tM}\right)\Big|_{t=0} \\ &= \pi_1 \left(1 - \frac{p_-}{(p_+ - p_-)^2}\right) \end{aligned} \tag{4.4.13}$$

and

$$\begin{aligned} E(M^2) &= \frac{\partial^2}{\partial t^2} E\left(e^{tM}\right)\Big|_{t=0} \\ &= \pi_1 \left(1 + \frac{2p_-^2}{(p_+ - p_-)^3} + \frac{p_-}{(p_+ - p_-)^2}\right) \end{aligned} \tag{4.4.14}$$

It is easy to check the mean-variance relationship based on the moments of the equilibrium distribution. We have,

$$\begin{aligned} &Var(M) - E(M) \\ &= \frac{-p_+^4 - p_-^4 + 2\,p_+p_- + 4\,p_+^2p_- - 6\,p_+p_-^2 + 4\,p_+^3p_- - 6\,p_+^2p_-^2 + 4\,p_+p_-^3 - p_-^2 + 2\,p_-^3}{(p_+ - p_-)^2(1 + p_+ - p_-)^2} \\ &= \frac{2\,p_+p_-\left(2\,(p_+ - p_-)^2 + 2\,p_+p_- + 2\,p_+ - 3\,p_- + 1\right) - (p_+^2 + p_-^2)^2 - p_-^2(1 - 2\,p_-)}{(p_+ - p_-)^2(1 + p_+ - p_-)^2} \end{aligned}$$

Figure 4.4 plots the function above.

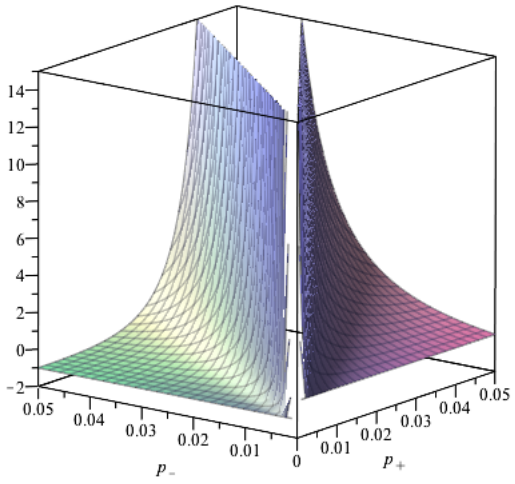


Figure 4.4: *The "variance minus mean" function based on the moments derived from the equilibrium distribution of the Markov chain.*

Notice that although the difference $Var(M) - E(M)$ can be negative for some parameter values (see Fig. 4.4), for the majority of the parameter space, we have over-dispersion i.e., $Var(M) > E(M)$.

4.5 The effect of time-scale

The importance of time scale in the analysis of neural spike trains was discussed in detail in Chapter 2. Employing the Skellam process with resetting for the analysis of neural spike trains, it is interesting to study the time-scale-dependent properties of this model.

4.5.1 Interspike interval distribution

We start studying the interspike interval in discrete time and address the continuous time as the limiting case. Consider $X_i \in \{-1, 0, 1\}$ defined in Equation (4.3.4). Then, $S_n = \sum_{i=1}^{n} X_i$ forms a random walk with state space $\{-1, 0, +1\}$. Given that $S_0 = 0$, we are interested in the first passage time to $+1$. Define the random variable $N_0 = \min\{n : S_n = +1\}$. Since the starting point is 0 and spike times are associated with $S(t) = +1$, the discrete version of interspike interval (ISI) is, in fact, $N = N_0 + 1$ (this is due to the extra 0 imposed by resetting). We assume $\lambda_1 > \lambda_2$ so that $\Pr(N < \infty) = 1$. The probability generating function (pgf) of N is

$$\Phi(t) = \frac{(1 - tp_0) - \sqrt{(1 - tp_0)^2 - 4t^2 p_- p_+}}{2p_-}. \tag{4.5.15}$$

The derivation of this generating function, Equation (4.5.15), is provided in Section 4.8.

Calculating the moments of the interspike interval $(N + 1)$ based on the derivatives of the probability generating function (4.5.15), we have

$$\begin{aligned} E(N) &= \frac{\partial}{\partial t}\Phi(t)\Big|_{t=1} \\ &= 1 + \frac{1}{p_+ - p_-}, \end{aligned} \tag{4.5.16}$$

$$\begin{aligned} E(N^2) &= \frac{\partial^2}{\partial t^2}\Phi(t)\Big|_{t=1} + E(N) \\ &= \frac{2p_+}{(p_+ - p_-)^3} + \frac{1}{p_+ - p_-} + 1\,. \end{aligned} \tag{4.5.17}$$

These equations allow for the moment estimates of the parameters p_- and p_+ (equivalently λ_1 and λ_2). We have

$$\begin{aligned}\widehat{p}_+ &= \frac{\overline{N^2}-\overline{N}}{2(\overline{N}-1)^3}\\ \widehat{p}_- &= \frac{\overline{N^2}-\overline{N}}{2(\overline{N}-1)^3}-\frac{1}{\overline{N}-1}\end{aligned}$$

Based on the discrete time Interspike interval N, the continuous-time ISI is $T = Nh$. Replacing this in the equations above, dividing both sides by h, and taking the limit as $h \to 0$ we get

$$\begin{aligned}\widehat{\lambda}_1 &= \frac{\overline{T^2}}{2(\overline{T})^3}\\ \widehat{\lambda}_2 &= \frac{\overline{T^2}}{2(\overline{T})^3}-\frac{1}{\overline{T}}\end{aligned}$$

Performing the same process in Equations (4.5.16) and (4.5.17) we have

$$E(T) = \frac{1}{\lambda_1-\lambda_2} \tag{4.5.18}$$

$$E(T^2) = \frac{2\lambda_1}{(\lambda_1-\lambda_2)^3} \tag{4.5.19}$$

Notice that under Poisson process ($\lambda_2 = 0$), these formulas simplify to the first two moments of the exponential distribution. The parameter λ_2 contributes to the tail of the interspike interval, and makes it heavier relative to the exponential distribution of the Poisson process.

4.5.2 Connection with Brownian motion

Recall the random walk defined by the partial sums $S_n = \sum_{i=1}^n X_i$ where X_i (i.i.d.) is defined in Equation (4.3.4). For any $m \geq 1$, and for any positive sequence $\{t_1, ..., t_m\}$, $t_i \neq t_j$ for $i \neq j$, the multivariate C.L.T., (Anderson, 2003, p.86), and the Kolmogorov's existence theorem, (Øksendal, 2010, p.10), imply that all finite dimensional distributions of the process $\{\frac{S_{\lfloor t/h \rfloor}}{\sqrt{\lfloor t/h \rfloor}}, t > 0\}$ converge to multivariate normal distribution as $h \to 0$ (or equivalently, as $n = h^{-1} \to \infty$). The mean and the variance of X_i are

$$\begin{aligned}\mu_x &= p_+ - p_-\,,\\ \sigma_x^2 &= (p_+ + p_-) - (p_+ - p_-)^2\,.\end{aligned}$$

Employing Donsker's Theorem, Billingsley (1968), the process

$$S_n(t) = \frac{1}{\sigma_x\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}(X_i - \mu_x) + \big(nt - \lfloor nt\rfloor\big)\frac{1}{\sigma_x\sqrt{n}}\big(X_{\lfloor nt\rfloor+1} - \mu_x\big) \tag{4.5.20}$$

converges to a Wiener process. Therefore, $B_n(t) = (\lambda_1 - \lambda_2)t + \sqrt{\lambda_1 + \lambda_2}S_n(t)$ converges to a Brownian motion (call it B_t) with drift $\mu = \lambda_1 - \lambda_2$ and diffusion parameter $\sigma = \sqrt{\lambda_1 + \lambda_2}$. Equivalently, one can say that the process $\{\frac{S_{\lfloor t/h\rfloor}}{\sqrt{\lfloor t/h\rfloor}}, t > 0\}$ converges to the Brownian motion B_t with drift $\mu = \lambda_1 - \lambda_2$ and diffusion parameter $\sigma = \sqrt{\lambda_1 + \lambda_2}$ (with the interpolation for continuity correction of the sample path).

It is important to recall that the difference between the two Poisson processes N_1 and N_2, which represent the counting process version of the inhibitory and excitatory synaptic input, form the random walk S_n discussed above. This random walk does, in a way, play the role of the membrane potential, which in long run, is a Brownian excursion.

4.5.3 Large time-scales

Now let us consider time scales much larger than the interspike intervals. What constitutes "large" depends on parameters λ_1 and λ_2. In particular, if N_T denotes the spike count in a time widow of size T, we are interested in time windows during which $\Pr(N_T \leq 0|\lambda_1, \lambda_2) < \epsilon$, for a given $\epsilon > 0$. For example, the area on or under each curve in Figure 4.5 shows the parameter space which satisfies $\Pr(N_T \leq 0|\lambda_1, \lambda_2) < \epsilon$, for $\epsilon \in \{0.001, 0.005, 0.01, 0.05\}$.

Now, consider a Skellam process S_0 with parameters (λ_1, λ_2) within the aforementioned parameter space. Modelling neural spike within the framework of Skellam process with resetting S, the spike times are $\{t : S(t) = 1\}$. Notice that these spike times are also the record times of the underlying Skellam process S_0 from which the SPR S has been derived. Based on the Skellam process S, and spike times $t_1, t_2, \cdots t_{N_T}$, the number of spikes in the time interval $(0, T]$, N_T, can be written as

$$N_T = S_{t_1} + (S_{t_2} - S_{t_1}) + (S_{t_3} - S_{t_2}) + \cdots + (S_{t_{N_T}} - S_{t_{N_T-1}}), \tag{4.5.21}$$

which is clearly a Skellam random variable with parameters $\lambda_1 t_{N_T}$ and $\lambda_2 t_{N_T}$ as each term in the summation above is a Skellam random variable. Figure 4.6 show a simulation study where a Skellam process with parameters $(21, 5)$ has been simulated 500 times. Assuming $\epsilon = 0.001$, $\Pr(N_T \leq 0|\lambda_1 = 21, \lambda_2 = 5) < \epsilon$. The histogram is based

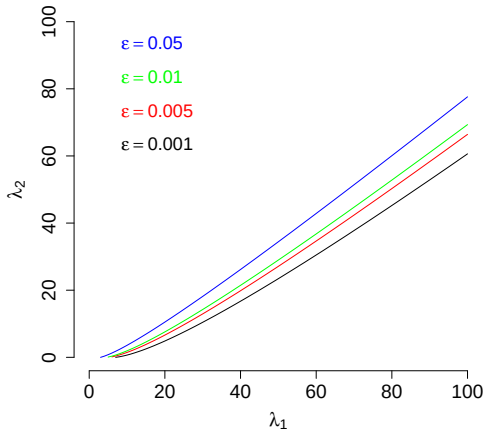


Figure 4.5: *The area on or under each curve shows the parameter space in which* $Pr(N_T \leq 0 \,|\, \lambda_1, \lambda_2) < \epsilon$.

on the spike counts over the observation interval, and the red curve is the Skellam density function with the same parameters to those used to simulate the Skellam process.

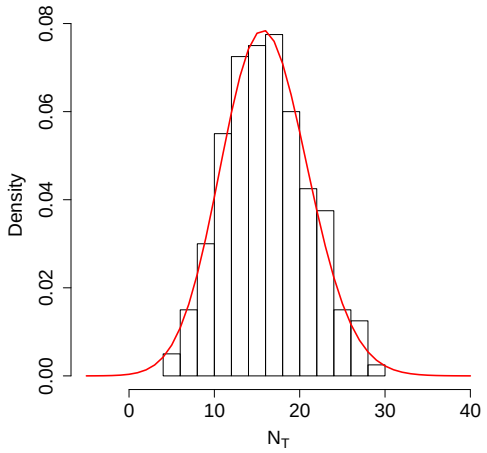


Figure 4.6: *The histogram of the simulated spike counts from a Skellam process with parameters* $(\lambda_1, \lambda_2) = (21, 5)$. *The superimposed red curve is the Skellam density function with the same parameters.*

4.6 Spike count and ISI variability

Let $N(t)$ show the number of spikes up to and including time t. Conventional models for $N(t)$ (usually Poisson) suffer from lack of flexibility in their mean-variance structures. While Poisson variability assumption is valid in some cases, significant departure from Poisson variability is usually observed in different areas of the brain, Maimon and Assad (2009); Nelson (2002), or in different time-scales, Ramezan et al. (2014). This departure from Poisson variability can also be due to the type of stimulus signal to which the neuron has been exposed, de Ruyter van Steveninck et al. (1997). We showed that in large time scales, the spike count in SPR follows a Skellam distribution, hence is an over-dispersed model (relative to Poisson). Recall that under this framework $E(N(t)) = \lambda_1 - \lambda_2$ and $Var(N(t)) = \lambda_1 + \lambda_2$. Figure 4.7 shows a simulation study which confirms this property of SPR.

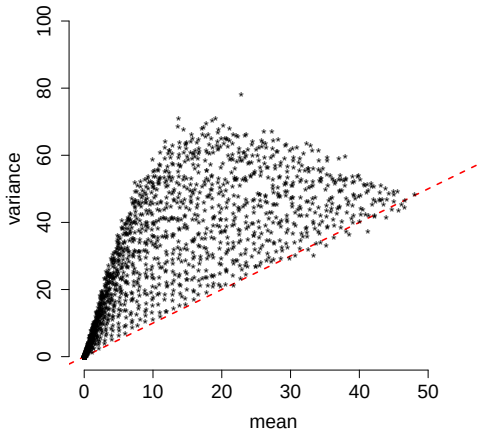


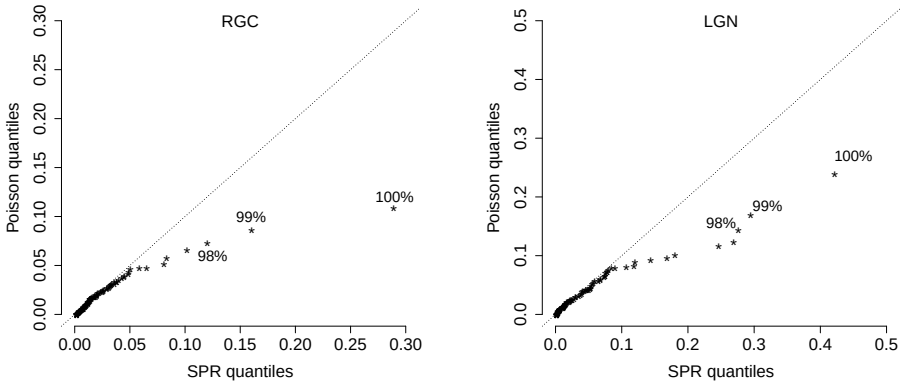
Figure 4.7: *Mean-variance relationship in the Skellam process with resetting. Each point in the plot represents the mean and variance of the spike counts for 1000 trials where* $\lambda_1, \lambda_2 \in \{1, 2, ..., 50\}$. *The red dashed line represents mean=variance.*

We can also look at the dispersion problem from an ISI point of view. The coefficient of variation of the ISI distribution (CV_{ISI}) is a measure of dispersion. For a Poisson model, $CV_{ISI} = 1$. Modelling ISI with Gamma or Inverse Gamma distributions, $CV_{ISI} < 1$, and for certain parameter values Inverse Gaussian can result in $CV_{ISI} > 1$. Based on the first two moments of the ISI in the Skellam model,

$$CV_{ISI} = \sqrt{\frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2}} > 1.$$

The retinogeniculate synapse data which was discussed in Chapter 3 is overdispersed relative to Poisson model, and has $\widehat{CV}_{ISI} > 1$. This makes the Gamma and the Inverse-Gamma distributions inappropriate for such data, but clearly, the ISI distribution based on the Skellam model is more approperiate.

From the retinogeniculate synapse data, we have randomly chosen a trial and fitted the SPR as well as Poisson process model. Figure 4.8 shows the ISI percentiles of simulated data from these two models, which confirms the heavier tail of SPR. Each dot in this plot shows a percentile.



(a) (b)

Figure 4.8: *The percentiles of the interspike intervals of Poisson process and SPR (in seconds). Panel (a) represents percentiles based on fitted models to an RGC trial and panel (b) shows those of an LGN trial. The numbers on the dots represents the percentiles, i.e.* 100^{th}, 99^{th}, 98^{th},... percentiles.

Table 4.1 includes the first two sample moments based on the real data and simulated SPR and Poisson models. It is clear that SPR values are closer to real data, particularly for the second moment.

Table 4.1: *The first two moments of the interspike intervals (ISI) based on real data, and simulated Poisson and SPR data.*

Neuron	**Data**	$\widehat{\mathbf{E}}(\mathbf{ISI})$	$\widehat{\mathbf{E}}(\mathbf{ISI^2})$
RGC	Real	0.0202	0.00041
	SPR	0.0215	0.00046
	Poisson	0.0192	0.00071
LGN	Real	0.0371	0.0014
	SPR	0.0394	0.0016
	Poisson	0.035	0.0028

4.7 Discussion

This chapter introduced and studied the properties of univariate Skellam process and univariate Skellam process with resetting (SPR) in both homogeneous and inhomogeneous cases. We also discussed the theoretical properties of the Skellam model under different time scales.

One motivation for utilizing Skellam process in the analysis of neural spike trains is that it allows for the over-dispersed spike counts (relative to Poisson), which has been previously addressed by using Negative Binomial models, Onken et al. (2009); Pillow and Scott (2012). However, multivariate extensions of Negative Binomial models do not have flexible dependency structure for the simultaneous analysis of multiple neurons. In Chapter 6 we introduce the multivariate version of SPR, which is capable of modelling multiple neural spike trains in a multivariate point process frame-work. The other advantage of SPR over the conventional methods is that it has nice biological interpretation, where the Skellam process is motivated by neural integration, and the resettings are due to the refractoriness of neurons.

It is important to emphasize that we have assumed that the two Poisson processes N_1 and N_2, which define the Skellam process $\{S_0(t) = N_1(t) - N_2(t),\, t \geq 0\}$, are independent. We also mentioned that these two processes *mimic* the integrated excitatory and the inhibitory activity of presynaptic neurons. The spiking activities of the presynaptic neurons are not necessarily independent and this is why we only used the neural integration as an analogy to Skellam process, furthermore, the neural integration is not simply the linear process of adding up the excitatory and inhibitory effects.

The main body of SPR is formed by the records of Skellam process. The resettings, essentially, delay the records by some values $h > 0$. Unfortunately, there is not much literature on the records of relatively more complex processes such as Skellam process. Studying the records of Skellam process is of both mathematical and biological interest. Our model is suitable for the ISIs because of its relatively long tail. More detailed theoretical study on the records of the Skellam process will help developing better models for interspike intervals.

Skellam process with resetting (SPR) introduced in this chapter can be generalized to the multivariate case in order to analyze multiple neural spike trains simultaneously. This multivariate problem is discussed in the Chapter 6.

4.8 Proofs of the theorems and some preliminary results on the mean-variance relationship

This appendix provides the proofs to the theorems discussed in this chapter, as well as some preliminary results on the mean-variance relationship in the Skellam process with resetting.

4.8.1 Proof of Theorem 4.1

Proof: First, we show that Definition 4.1 implies Definition 4.2. Let $\{S(t), t \geq 0\}$ be a Skellam process. Therefore, $\{S(t), t \geq 0\}$ satisfies the conditions listed in Definition 4.1. Following a similar proof to that of (Grimmett and D., 2001, pp.247-248), we show that for any values $s, t \geq 0$, the probability generating function of the random variable $S(t+s) - S(t)$ is the same as that of a Skellam random variable with nonnegative parameters $(\lambda_1 t, \lambda_2 t)$. Let us show define $S^*(t,s) = S(t+s) - S(t)$

$$\begin{aligned}\Pr\left(S^*(t+h,s) = j\right) &= \sum_{i=-\infty}^{\infty} \Big\{ \Pr\big[S^*(t,s) = i\big] \Pr\big[S^*(t+h,s) = j | S^*(t,s) = i\big] \Big\} \\ &= \sum_{i=j-1}^{j+1} \Big\{ \Pr\big[S^*(t,s) = i\big] \Pr\big[S^*(t+h,s) = j | S^*(t,s) = i\big] \Big\} + o(h).\end{aligned}$$

Define the notation $P_j(t,s) = \Pr\left[S^*(t,s) = j\right]$. We have,

$$P_j(t+h,s) = \lambda_1 h P_{j-1}(t,s) + P_j(t,s) - (\lambda_1 + \lambda_2) h P_j(t,s) + \lambda_2 h P_{j+1}(t,s) + o(h).$$

Taking $P_j(t,s)$ to the other side, dividing by h and taking the limit as $h \to 0$ we have

$$\frac{d}{dt}P_j(t,s) = \lambda_2\Big(P_{j+1}(t,s) - P_j(t,s)\Big) - \lambda_1\Big(P_j(t,s) - P_{j-1}(t,s)\Big). \quad (4.8.22)$$

Equation (4.8.22) is a differential-difference equation for $P_j(t,s)$. The boundary condition is $P_j(t,t) = P_j(s,s) = I_{\{0\}}(j)$ where $I_{\{0\}}(j)$ is an indicator function. Let $G(w,t,s) = \sum_{j=-\infty}^{\infty} P_j(t,s)w^j$ denote the probability generating function of $S^*(t,s)$. Clearly, $G(0,t,s) = 0$. For $w \neq 0$ we have,

$$\begin{aligned}
\sum_{j=-\infty}^{\infty} \frac{d}{dt}P_j(t,s)w^j &= \sum_{j=-\infty}^{\infty} \Big\{\lambda_2\Big(P_{j+1}(t,s) - P_j(t,s)\Big) - \lambda_1\Big(P_j(t,s) - P_{j-1}(t,s)\Big)\Big\}w^j \\
&= \lambda_2\Big[\frac{1}{w}\sum_{j=-\infty}^{\infty} P_{j+1}(t,s)w^{j+1} - \sum_{j=-\infty}^{\infty} P_j(t,s)w^j\Big] \\
&\quad -\lambda_1\Big[\sum_{j=-\infty}^{\infty} P_j(t,s)w^j - w\sum_{j=-\infty}^{\infty} P_{j-1}(t,s)w^{j-1}\Big] \\
&= \lambda_2\Big[\frac{1}{w}G(w,t,s) - G(w,t,s)\Big] - \lambda_1\Big[G(w,t,s) - wG(w,t,s)\Big]
\end{aligned}$$

so,

$$\frac{\partial G(w,t,s)}{\partial t} = G(w,t,s)\Big[-(\lambda_1+\lambda_2) + \lambda_1 w + \frac{\lambda_2}{w}\Big] \quad (4.8.23)$$

with the boundary condition $G(w,t,t) = G(w,s,s) = 1$. Integrating both sides, the unique solution to the differential equation in (4.8.23) is,

$$G(w,t,s) = \exp\Big(-(\lambda_1+\lambda_2)t + \lambda_1 t w + \frac{\lambda_2 t}{w}\Big),$$

which is the probability generating function of a Skellam distribution with parameters $\lambda_1 t$ and $\lambda_2 t$, Equation (4.2.2).

We now prove the other direction, i.e., Definition 4.2 implies Definition 4.1. It suffices to show that for any value $t \geq 0$ and a positive infinitesimal h conditions 3-6 of Definition 4.1 are met. Writing the Taylor expansions, we prove that condition 3 holds true. The proofs for the other conditions are similar.

$$\begin{aligned}
\Pr\Big[S(t+h) - S(t) = 0\Big] &= \exp\{-(\lambda_1+\lambda_2)h\}\sum_{y=0}^{\infty}\frac{(\lambda_1 h)^y(\lambda_2 h)^y}{(y!)^2} \\
&= \Big(1 - (\lambda_1+\lambda_2)h + o(h)\Big)\Big(1 + o(h)\Big) \\
&= 1 - (\lambda_1+\lambda_2)h + o(h) \quad (4.8.24)
\end{aligned}$$

Notice that in the second line of the equations above we used the fact that

$$1 < \sum_{y=0}^{\infty} \frac{(\lambda_1 h)^y (\lambda_2 h)^y}{(y!)^2} \leq \exp\{\lambda_1 \lambda_2 h^2\} = 1 + o(h)$$

4.8.2 Proof of Theorem 4.2

Proof: N_1 and N_2 are Poisson processes, so $N_1(0) = N_2(0) = 0$, and consequently $S(0) = 0$. For all values $t_1 < t_2$, $N_1(t_2) - N_1(t_1)$ is independent of the times of incidents of N_1 during the interval $[0, t_1]$. The same property holds true for N_2. Therefore, $S(t_2) - S(t_1)$ is independent of the times of incidents of S during the interval $[0, t_1]$. Notice that the times of incidents of S is the union of those of N_1 and N_2. The last piece to complete the proof is to show that for any $s, t \geq 0$, $S(t+s) - S(s) \sim Sk(\lambda_1 t, \lambda_2 t)$. Notice that

$$S(t+s) - S(s) \quad = \quad \Big[N_1(t+s) - N_1(s)\Big] - \Big[N_2(t+s) - N_2(s)\Big].$$

However,

$$\begin{aligned} N_1(t+s) - N_1(s) &\sim \text{Pois}(\lambda_1 t), \\ N_2(t+s) - N_2(s) &\sim \text{Pois}(\lambda_2 t), \end{aligned}$$

and since $\{N_1, t \geq 0\}$ and $\{N_2, t \geq 0\}$ are independent,

$$S(t+s) - S(s) \sim Sk(\lambda_1 t, \lambda_2 t).$$

4.8.3 Proof of Theorem 4.3

Proof: First, we show that Definition 4.3 implies Definition 4.4. Let $\{S(t), t \geq 0\}$ be a Skellam process with regards to Definition 4.3. Notice that

$$S(t+s) - S(s) \sim Sk\Big(\int_s^{t+s} \lambda_1(y)\, dy \ , \ \int_s^{t+s} \lambda_2(y)\, dy\Big),$$

can be written as

$$S(t+s) - S(s) \sim Sk\Big(\int_0^{t} \lambda_1(y+s)\, dy \ , \ \int_0^{t} \lambda_2(y+s)\, dy\Big), \tag{4.8.25}$$

by assuming that the inhomogeneous Skellam process starts at time s. Let us define $S^*(t, s) = S(t+s) - S(s)$. Similar to the proof of Theorem 4.1 we have

$$\begin{aligned} P_j(t+s+h, s) \quad = \quad & \lambda_1(t+s) h P_{j-1}(t+s, s) + P_j(t+s, s) \\ & - \Big(\lambda_1(t+s) + \lambda_2(t+s)\Big) h P_j(t+s, s) \\ & + \lambda_2(t+s) h P_{j+1}(t+s, s) + o(h), \end{aligned}$$

where $P_j(t+s,s) = \Pr\left[S^*(t+s,s) = j\right]$. Rearranging the terms, dividing both sides by h and taking the limit as h goes to zero we have

$$\begin{aligned}\frac{d}{dt}P_j(t+s,s) &= \lambda_2(t+s)\Big(P_{j+1}(t+s,s) - P_j(t+s,s)\Big) \\ &\quad -\lambda_1(t+s)\Big(P_j(t+s,s) - P_{j-1}(t+s,s)\Big).\end{aligned}$$

which is a differential-difference equation similar to Equation (4.8.22). Following the same approach we get,

$$\begin{aligned}\frac{\partial G(w,t+s,s)}{\partial t} &= G(w,t+s,s)\Big[-\Big(\lambda_1(t+s)+\lambda_2(t+s)\Big) \\ &\quad +\lambda_1(t+s)w + \frac{\lambda_2(t+s)}{w}\Big],\end{aligned}$$

with the boundary condition $G(w,t,t) = G(w,s,s) = 1$. Integrating both sides, the unique solution to this equations is,

$$\begin{aligned}G(w,t+s,s) &= \exp\left\{-\Big(\int_0^t \lambda_1(y+s)\,dy + \int_0^t \lambda_2(y+s)\,dy\Big)\right. \\ &\quad \left. +w\int_0^t \lambda_1(y+s)\,dy + \frac{\int_0^t \lambda_1(y+s)\,dy}{w}\right\},\end{aligned}$$

which is the probability generating function of a the Skellam random variable defined in Equation (4.8.25).

The equivalence of the two definitions is complete if it is proved that Definition 4.4 implies Definition 4.3. According to Definition 4.4,

$$\begin{aligned}\Pr\Big[S(t+h) - S(t) = i\Big] &= \exp\left\{-\left(\int_t^{t+h}\lambda_1(y)\,dy + \int_t^{t+h}\lambda_2(y)\,dy\right)\right\} \times \\ &\quad \sum_{y=0}^{\infty}\frac{\left(\int_t^{t+h}\lambda_1(y)\,dy\right)^{y+i}\left(\int_t^{t+h}\lambda_2(y)\,dy\right)^{y}}{(y+i)!\,y!} \qquad (4.8.26)\end{aligned}$$

For $i = 1, 2$, write

$$m_h^{(i)}(t) = \int_t^{t+h}\lambda_i(y)\,dy$$

Writing the Taylor expansion of $m_h^{(i)}(t)$, $i = 1, 2$ with respect to h about $h = 0$ we have,

$$m_h^{(i)}(t) = \lambda_i(t)h + o(h) \qquad (4.8.27)$$

Substituting Equation (4.8.27) in Equation (4.8.26) we get,

$$\begin{aligned}\Pr\Big[S(t+h)-S(t)=i\Big] &= \exp\Big\{-\Big(\lambda_1(t)+\lambda_2(t)\Big)h+o(h)\Big\}\times \\ &\quad \sum_{y=0}^{\infty}\frac{\Big(\lambda_1(t)h\Big)^{y+i}\Big(\lambda_2(t)h\Big)^{y}}{(y+i)!\,y!}\end{aligned}$$

Now, writing the Taylor expansion of this probability function and following the same argument as that of Equation (4.8.24) for conditions 1-6, the proof is complete.

An alternative proof to the second part of the proof above is to write Equation (4.8.26) as

$$\Pr\Big[S(t+h)-S(t)=i\Big] = \Pr\Big[N_1(t+h)-N_1(t)-\big(N_2(t+h)-N_2(t)\big)=i\Big],$$

and calculate it based on the value of i. For example if $i=0$,

$$\begin{aligned}\Pr\Big[N_1(t+h)-N_1(t)-\big(N_2(t+h)-N_2(t)\big)=0\Big] &= \Pr\Big[N_1(t+h)-N_1(t)=0\Big]\times \\ &\quad \Pr\Big[\big(N_2(t+h)-N_2(t)\big)=0\Big]+o(h) \\ &= 1-\lambda_1(t)\,h-\lambda_2(t)\,h+o(h)\end{aligned}$$

The calculations based on other values of i are similar.

4.8.4 Proof of Equation (4.5.15)

Conditioning on the first step, the probability generating function of N can be derived as follows.

$$\begin{aligned}\Phi(t) &= E(t^{N_0}) \\ &= E\left(t^{N_0}\middle|X_1=1\right)\,p_+ + E\left(t^{N_0}\middle|X_1=0\right)\,p_0 + E\left(t^{N_0}\middle|X_1=-1\right)\,p_- \qquad (4.8.28)\end{aligned}$$

Given $X_1=1$, we have $N_0=1$, and the first term in the R.H.S. is $t\,p_+$. If $X_1=-1$, define N_1 and N_2 to be, respectively, the number of steps required to go from -1 to 0, and the number of steps required to go from 0 to $+1$. Clearly, $N_0=1+N_1+N_2$. Since N_1 and N_2 depend on different subsets of X_is, they are independent. Furthermore, N_1 and N_2 are both first passage times, so they are equally distributed as the random variable N_0. Therefore,

$$\begin{aligned} E\left(t^{N_0}\middle|X_1=-1\right) &= E\left(t^{1+N_1+N_2}\middle|X_1=-1\right) \\ &= t\,E(t^{N_1})\,E(t^{N_2}) \\ &= t\left(\Phi(t)\right)^2. \end{aligned} \tag{4.8.29}$$

Similarly, given $X_1 = 0$ we have $N_0 = 1 + N_2$. Therefore,

$$\begin{aligned} E\left(t^{N_0}\middle|X_1=0\right) &= E\left(t^{1+N_2}\middle|X_1=0\right) \\ &= t\,\Phi(t). \end{aligned} \tag{4.8.30}$$

Substituting Equations (4.8.29) and (4.8.30) in Equation (4.8.28) we get

$$tp_{-}\left(\Phi(t)\right)^2 + (tp_0 - 1)\Phi(t) + tp_{+} = 0.$$

Solving this quadratic equation of $\Phi(t)$ based on the initial condition $\Phi(0) = 0$, the valid solution is

$$\Phi(t) = \frac{(1-tp_0) - \sqrt{(1-tp_0)^2 - 4t^2p_{-}p_{+}}}{2tp_{-}}. \tag{4.8.31}$$

Multiplying this equation by t (to get the pgf of $N = N_0 + 1$) completes the proof.

4.8.5 Some preliminary results on the mean-variance relationship in SPR

Here we discuss some of the mathematical properties of Skellam process with Resetting.

Lemma 4.1 *Let $N(t)$ be the random variable representing the number of records of a Skellam process $S_0 = \{S_0(t), t \geq 0\}$ up to and including time t, and let $\{S(t), t \geq 0\}$ be the SPR generated from the process S_0. Then*

$$\begin{aligned} Pr\Big(N(t)=n\Big) &= \lambda_1 \int_0^t \Big[Pr\big(N(u)=n-1|S(u)=0\big) - \\ &\qquad Pr\big(N(u)=n-1|S(u)=0\big)\Big] Pr\big(S(u)=0\big)\,du \end{aligned}$$

where $Pr(N(0)=0)=1$.

Proof: Conditioning on $N(t)$ and $S(t)$ we have,

$$\begin{aligned}
\Pr\Big(N(t+h)=n\Big) &= p_{+}\Pr\Big(S(t)=0\Big|N(t)=n-1\Big)\Pr\Big(N(t)=n-1\Big) \\
&\quad +\Pr\Big(S(t)=1\Big|N(t)=n\Big)\Pr\Big(N(t)=n\Big) \\
&\quad +\Pr\Big(S(t)=0\Big|N(t)=n\Big)\Pr\Big(N(t)=n\Big) \\
&\quad -p_{+}\Pr\Big(S(t)=0\Big|N(t)=n\Big)\Pr\Big(N(t)=n\Big) \\
&\quad +\Pr\Big(S(t)<0\Big|N(t)=n\Big)\Pr\Big(N(t)=n\Big) \\
&= \Pr\Big(N(t)=n\Big)+p_{+}\Big[\Pr\Big(S(t)=0\Big|N(t)=n-1\Big)\Pr\Big(N(t)=n-1\Big) \\
&\quad -\Pr\Big(S(t)=0\Big|N(t)=n\Big)\Pr\Big(N(t)=n\Big)\Big].
\end{aligned}$$

Taking $\Pr\big(N(t)=n\big)$ to the other side, dividing by h and taking the limit as h goes to zero we have,

$$\begin{aligned}
\frac{d}{dt}\Pr\Big(N(t)=n\Big) &= \lambda_1\Big[\Pr\Big(N(t)=n-1\Big|S(t)=0\Big) \\
&\quad -\Pr\Big(N(t)=n\Big|S(t)=0\Big)\Big]\Pr\Big(S(t)=0\Big). \qquad (4.8.32)
\end{aligned}$$

Integrating both sides of Equation (4.8.32) with respect to t we have

$$\begin{aligned}
\Pr\Big(N(t)=n\Big) &= c+\lambda_1\int_0^t\Big[\Pr\Big(N(u)=n-1\Big|S(u)=0\Big) \\
&\quad -\Pr\Big(N(u)=n\Big|S(u)=0\Big)\Big]\Pr\Big(S(u)=0\Big)\,du, \qquad (4.8.33)
\end{aligned}$$

where

$$\Pr\Big(N(t)=n\Big|S(t)=0\Big)=0 \quad \text{for } n<0.$$

The initial condition

$$\Pr\Big(N(0)=0\Big)=1$$

implies that $c=0$ is the unique solution to Equation (4.8.33).

Lemma 4.2 *Consider the assumptions of Lemma 4.1. Furthermore, assume that $t<\infty$ and $Pr(S(t)=0)<1$. Then all of the followings are finite;*

(i) $E\Big(N(t)\Big)$

(ii) $E\Big([N(t)]^2\Big)$

(iii) $E\Big(N(t)\Big|S(t)=0\Big)$

(iv) $E\Big([N(t)]^2\Big|S(t)=0\Big)$

Proof: Let $\{N_i(t), t \geq 0\}$ $i=1,2$, be two Poisson processes constructing the Skellam process $\{S_0(t) = N_1(t) - N_2(t), t \geq 0\}$, which itself generates the SPR $\{S(t), t \geq 0\}$. To prove (i) notice that for $t < \infty$

$$N(t) \leq N_1(t) + N_2(t). \tag{4.8.34}$$

In other words, the number of records of the process S during the period $(0,t]$ can not be more than the aggregate number of events from the two processes N_1 and N_2. Therefore

$$E\Big(N(t)\Big) \leq E\Big(N_1(t)\Big) + E\Big(N_2(t)\Big) < \infty.$$

Similarly, to prove (ii), square the two sides of the inequality (4.8.34) and then take expectation from both sides. For (iii) we have

$$\begin{aligned}
E\Big(N(t)\Big|S(t)=0\Big) &= \sum_{n=0}^{\infty} n \Pr\Big(N(t)=n\Big|S(t)=0\Big) \\
&= \sum_{n=0}^{\infty} n \frac{\Pr\Big(N(t)=n\Big)\Pr\Big(S(t)=0\Big|N(t)=n\Big)}{\Pr\Big(S(t)=0\Big)} \\
&\leq \frac{1}{\Pr\Big(S(t)=0\Big)} \sum_{n=0}^{\infty} n \Pr\Big(N(t)=n\Big) \\
&= \frac{1}{\Pr\Big(S(t)=0\Big)} E\Big(N(t)\Big) \\
&< \infty
\end{aligned}$$

The proof of (iv) is similar to (iii).

Theorem 4.5 *Consider the assumptions of Lemma 4.2. The first two moments of $N(t)$, are*

$$\begin{aligned}
E\Big(N(t)\Big) &= \lambda_1 \int_0^t Pr\Big(S(u)=0\Big)\, du \qquad (4.8.35) \\
E\Big([N(t)]^2\Big) &= 2\lambda_1 \int_0^t E\Big(N(u)\Big|S(u)=0\Big) Pr\Big(S(u)=0\Big)\, du + E\Big(N(t)\Big)
\end{aligned}$$

Proof: According to Lemma 4.2, $E(N(t))$ is finite. Employing Lemma 4.1 we have

$$\begin{aligned} E\Big(N(t)\Big) &= 0+\sum_{n=1}^{\infty} n\Pr\Big(N(t)=n\Big) \\ &= \sum_{n=1}^{\infty}\Bigg\{n\lambda_1\int_0^t\Big[\Pr\Big(N(u)=n-1\Big|S(u)=0\Big) \\ &\quad -\Pr\Big(N(u)=n\Big|S(u)=0\Big)\Big]\Pr\Big(S(u)=0\Big)\,du\Bigg\} \end{aligned}$$

where the integral is over a finite interval ($t\in(0,T]$) and the integrand is a bounded function of u. This along with using Lemma 4.2 allows us to switch the order of the integral and the summation and break the summation into two parts as follows

$$\begin{aligned} E\Big(N(t)\Big) &= \lambda_1\int_0^t\sum_{n=1}^{\infty} n\Pr\Big(N(u)=n-1\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &\quad -\lambda_1\int_0^t\sum_{n=1}^{\infty} n\Pr\Big(N(u)=n\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &= \lambda_1\int_0^t\Big[E\Big(N(u)\Big|S(u)=0\Big)+1\Big]\Pr\Big(S(u)=0\Big)\,du \\ &\quad -\lambda_1\int_0^t E\Big(N(u)\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &= \lambda_1\int_0^t\Pr\Big(S(u)=0\Big)\,du. \end{aligned}$$

Following similar argument to the previous part, $E[N(t)]^2$ is derived as follows,

$$\begin{aligned} E\Big([N(t)]^2\Big) &= 0+\sum_{n=1}^{\infty} n^2\Pr\Big(N(t)=n\Big) \\ &= \lambda_1\int_0^t\sum_{n=1}^{\infty} n^2\Pr\Big(N(u)=n-1\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &\quad -\lambda_1\int_0^t\sum_{n=1}^{\infty} n^2\Pr\Big(N(u)=n\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &= \lambda_1\int_0^t\Bigg[E\Big([N(u)]^2\Big|S(u)=0\Big)+ \\ &\quad \sum_{n=1}^{\infty}(2n-1)\Pr\Big(N(u)=n-1\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\Bigg]\,du \\ &\quad -\lambda_1\int_0^t E\Big([N(u)]^2\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \end{aligned}$$

$$
\begin{aligned}
&= \lambda_1 \int_0^t \left[\sum_{n=1}^{\infty} (n-1) \Pr\Big(N(u) = n-1 \Big| S(u) = 0\Big) \Pr\Big(S(u) = 0\Big) \right] du \\
&\quad + \lambda_1 \int_0^t \left[\sum_{n=1}^{\infty} n \Pr\Big(N(u) = n-1 \Big| S(u) = 0\Big) \Pr\Big(S(u) = 0\Big) \right] du \\
&= \lambda_1 \int_0^t E\Big(N(u) \Big| S(u) = 0\Big) \Pr\Big(S(u) = 0\Big)\, du \\
&\quad + \lambda_1 \int_0^t \Big[E\Big(N(u) \Big| S(u) = 0\Big) + 1 \Big] \Pr\Big(S(u) = 0\Big)\, du \\
&= 2\lambda_1 \int_0^t E\Big(N(u) \Big| S(u) = 0\Big) \Pr\Big(S(u) = 0\Big)\, du + E\Big(N(t)\Big)
\end{aligned}
$$

The formula for $E\big(N(t)\big)$, Equation (4.8.35), declares that the more likely the underlying Skellam process with resetting (SPR) is at zero, the higher the mean spike count. Clearly, increasing λ_2, which is the inhibition parameter, increases $\Pr\big(S(t) < 0\big)$, which consequently decreases the mean spike count $E\big(N(t)\big)$ as expected, and vice versa. As a special case, let us assume that $N = \{N(t), t \geq 0\}$ is a Poisson process with parameter λ_1. Therefore, every event in N is a record. We have,

$$
\begin{aligned}
\Pr\Big(S(t) = 0\Big) &= 1 - \Pr\Big(S(t) = 1\Big) \\
&= 1 - \lim_{h \searrow 0} \Pr\Big(\text{one event in } (t-h, t)\Big) \\
&= 1\,.
\end{aligned}
$$

Therefore, $E\big(N(t)\big) = \lambda_1 t$ as expected. Theorem 4.5 provides insight about the mean-variance behaviour of $N(t)$. Define

$$
g(t) = Var\Big(N(t)\Big) - E\Big(N(t)\Big).
$$

We have

$$
\begin{aligned}
g(t) &= E\Big([N(t)]^2\Big) - \left[E\Big(N(t)\Big)\right]^2 - E\Big(N(t)\Big) \\
&= 2\lambda_1 \int_0^t E\Big(N(u) \Big| S(u) = 0\Big) \Pr\Big(S(u) = 0\Big)\, du \\
&\quad - \left[\lambda_1 \int_0^t \Pr\Big(S(u) = 0\Big)\, du\right]^2 .
\end{aligned}
$$

However, if $V(t) = \lambda_1 \int_0^t \Pr\big(S(u) = 0\big)\, du$, then

$$
\Big(V(t)\Big)^2 = 2 \int_0^t \Big(\frac{d}{dw} V(w)\Big)\, V(w)\, dw\,.
$$

Thus,

$$\begin{aligned} g(t) &= 2\lambda_1 \int_0^t E\Big(N(u)\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &\quad -2\int_0^t \lambda_1 \Pr\Big(S(w)=0\Big)\Big[\lambda_1 \int_0^w \Pr\Big(S(u)=0\Big)\,du\Big]\,dw \\ &= 2\lambda_1 \int_0^t E\Big(N(u)\Big|S(u)=0\Big)\Pr\Big(S(u)=0\Big)\,du \\ &\quad -2\lambda_1 \int_0^t E\Big(N(w)\Big)\Pr\Big(S(w)=0\Big)\,dw \\ &= 2\lambda_1 \int_0^t \Big[E\Big(N(u)\Big|S(u)=0\Big) - E\Big(N(u)\Big)\Big]\Pr\Big(S(u)=0\Big)\,du \end{aligned}$$

with the boundary condition

$$g(t)\Big|_{t=0} = Var\Big(N(t)\Big) - E\Big(N(t)\Big)\Big|_{t=0} = 0.$$

Equation (4.8.36) states that the mean-variance relationship in SPR is a weighted average of the difference term

$$E\Big(N(t)\Big|S(t)=0\Big) - E\Big(N(t)\Big),$$

which is 0 in the special case of Poisson process. The departure from Poisson variability depends on the probability of the underlying SPR being at 0 over time. Clearly, if $\lambda_2 = 0$ or $\lambda_2 >> \lambda_1$, the probability term $\Pr\big(S(t)=0\big)$ and consequently $g(t)$ approaches zero as expected.

Chapter 5

Parameter Estimation in Univariate SPR

5.1 Introduction

In Chapter 4, the Skellam Process with Resetting (SPR) was introduced and its properties were studied. In this chapter, we derive the likelihood function based on the observable data (spike trains) and introduce a computationally efficient recursive algorithm for parameter estimation within SPR framework. In the time-inhomogeneous case, the multiscale estimation method of Chapter 2 is employed to estimate the functions $\lambda_1(t)$ and $\lambda_2(t)$ of the underlying SPR. A similar cross-validation scheme to that of Chapter 2 is introduced to choose the tuning parameters of the multiscale method properly. Simulation studies and the analyses of retinogeniculate synapse data provide promising results on the performance of the SPR.

5.2 Likelihood function

The likelihood function of the parameters based on the observable data (the observed spike trains) is derived in this section. Discretizing time into very small time bins, spike trains can be studied as realizations of a binary time series with values 1 and 0, where 1 and 0 show, respectively, the occurrence or lack of occurrence of a spike in that particular time bin. In the SPR setup, spikes are associated with the visits to state $\{1\}$, implying that the negative values of the Skellam Process with Resetting (see for illustration Figure 4.3) are not observable. In fact, these negative values are confounded with the zeros in spike train data. Therefore, the random variable X_i

defined in Equation (4.3.4) is not observable. Let X_i^{obs} show the value of the series at time bin i, Equation (4.3.3). We have

$$X_i^{obs} = \begin{cases} 1 & \text{if a spike occurs at time bin } i \\ 0 & \text{otherwise} \end{cases}$$

therefore, X_i^{obs} can also be formulated as follows,

$$X_i^{obs} = I\Big(\sum_{j<i}(X_j - X_j^{obs}) = 0\Big) \times I\Big(X_i = 1\Big), \tag{5.2.1}$$

where X_i is defined in Equation (4.3.4), $X_0 = X_0^{obs} = 0$, and

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{o.w.} \end{cases}$$

Notice that to have a spike at time i, two conditions would be met:

1. The value of the underlying Skellam process at time $i-1$, i.e. $\sum_{j<i} X_j$, should be non-negative. In other words, $\sum_{j<i}(X_j - X_j^{obs}) = 0$.

2. At time i, the value of the unobserved random variable X_i should be 1, showing the domination of inhibitory inputs, N_2, by the excitatory ones, N_1.

As mentioned in Chapter 4, for a given value of h limiting to 0, the probability function of X_i is

x_i	-1	0	1
$\Pr(X_i = x_i)$	p_-	p_0	p_+

where

$$p_- = \lambda_2 h + o(h), \tag{5.2.2}$$
$$p_0 = 1 - (\lambda_1 + \lambda_2)\, h + o(h), \tag{5.2.3}$$
$$p_+ = \lambda_1 h + o(h), \tag{5.2.4}$$

Interest lies in estimation of the two parameters λ_1 and λ_2, or equivalently p_- and p_+. The likelihood of these parameters based on $\boldsymbol{x} = (x_1, x_2, ..., x_k)$, which are k realizations of X_i, Equation (4.3.4), is

$$L(p_-, p_+; k_-, k_0, k_+) = \binom{k}{k_-, k_0, k_+}\, p_-^{k_-}\; p_0^{k_0}\; p_+^{k_+}, \tag{5.2.5}$$

where $k_- = \sum_j I(x_j = -1)$, $k_0 = \sum_j I(x_j = 0)$, $k_+ = \sum_j I(x_j = +1)$, $k_- + k_0 + k_+ = k$ and $p_- + p_0 + p_+ = 1$. Notice that we have used the independence assumption among

X_is to write this likelihood function. Clearly, the independence among $X_1, X_2, ...$ does not imply that X_i^{obs}s are independent of one another.

Equation (5.2.5) is not the likelihood function of the parameters based on the observable data (spike trains). Notice that the observable data associated with the small time bins are 0 (no spike), or 1 (occurrence of a spike). Therefore, during the intervals where no spikes are observed, the two values -1 and 0 and some of the $+1$s of the Skellam process with resetting are confounded, thus reported as 0 in the observable data. To better see this confounding, consider the observed spike train $(0, 0, 0, 0, 1)$. The actual path associated with this sequence could be $(0, -1, 1, 0, 1)$, or $(0, 0, -1, 1, 1)$, etc. In short, many possible vectors $\boldsymbol{x}$ can result in the same observed vector $\boldsymbol{x}^{obs}$. Figure 5.1 shows these possible paths for the above-mentioned example.

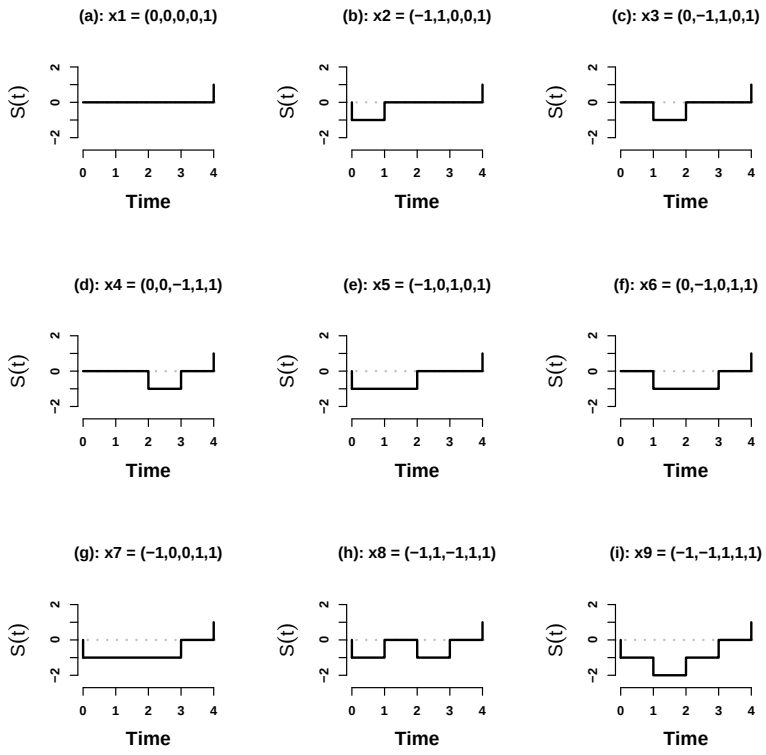


Figure 5.1: *A toy example for possible paths for* $\boldsymbol{x}^{obs} = (0, 0, 0, 0, 1)$. *Panel (a) is the only observable path while panels (b)-(i) show the different trajectories, all of which result in the same observed vector in panel (a).*

Based on only one discretized interspike interval $\boldsymbol{x}^{obs} = (0, 0, ..., 0, 1)$ of length $k + 1$,

the likelihood function of the parameters λ_1 and λ_2 is the sum of the likelihoods of all possible paths (see Figure 5.1),

$$L\Big(p_-, p_+; \boldsymbol{x^{obs}}\Big) = \sum_{i=1}^{\lceil k/2 \rceil} a_i \; p_-^{\alpha_i} \; p_0^{\beta_i} \; p_+^{\gamma_i}, \tag{5.2.6}$$

where p_-, p_0 and p_+ are defined, respectively, in Equations (5.2.2), (5.2.3) and (5.2.4) and a_i is the number of equiprobable paths with α_i number of -1s, β_i zeros and γ_i number of 1s. The length of the vector $\boldsymbol{x^{obs}} = (0, 0, ..., 0, 1)$ is $k+1$, which also shows the numbers of steps to a spike. Notice that steps are either -1, 0 or $+1$, however, due to resetting, the first step after each spike is always 0, forcing the process not to allow for two consecutive spikes in a short interval. Therefore, the "degrees of freedom" to choose steps equals to k. In this notation,

$$(\alpha_i, \beta_i, \gamma_i) \in \Big\{(\alpha, \beta, \gamma) : \; \alpha + \beta + \gamma = k, \; 0 \leq \alpha \leq \lfloor (k-1)/2 \rfloor, \; \gamma - \alpha = 1\Big\}.$$

Now, let us assume the starting point of the observation interval is at the first spike and the end of it is at the last spike. In Chapter 4 the modeling assumption of conditional independence of spike times was discussed. This assumption states that conditional on the current spike time, the past spike times are independent of the future ones, which is essentially, modeling the process as a Markov chain. This assumption has been addressed in the analysis of neural spike trains before, Kass and Ventura (2001). Recall that during the refractory period, most of the chemical mechanisms across the membrane and the membrane voltage reset, which provides biological justification for this conditional independence assumption. Based on these assumptions, the likelihood function of the parameters λ_1 and λ_2 given a spike train consists of n spikes is

$$L\Big(p_-, p_+; \boldsymbol{x}_1^{obs}, ..., \boldsymbol{x}_n^{obs}\Big) = \prod_{l=1}^{n-1} \sum_{i=1}^{\lceil (k_l-1)/2 \rceil} a_i \; p_-^{\alpha_i} \; p_0^{\beta_i} \; p_+^{\gamma_i}. \tag{5.2.7}$$

The product has $n-1$ terms as the time to the first spike is yet to be taken into consideration. The resettings are essential for the conditional independence, starting right after the first spike.

Table 5.1 illustrates the possible values for a_i, α_i, β_i, γ_i for a toy example when $\boldsymbol{x}^{obs} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$. Notice that here we assume that the first 0 is due to resetting.

It is noteworthy that the minimum value of β_i is one, which is because of the zeros occurring after each and every spike as a result of resetting. The parameters a_i can

Table 5.1: *Values associated with vectors* $\boldsymbol{x}$ *which are possible paths corresponding to the observable vector* $\boldsymbol{x}^{obs} = (0,0,0,0,0,0,0,0,0,0,0,1)$.

i	a_i	α_i	β_i	γ_i
1	1	0	11	1
2	55	1	9	2
3	660	2	7	3
4	2310	3	5	4
5	2310	4	3	5
6	462	5	1	6

be computed based on combinatorial calculations. Let n show the number of steps between two consecutive spikes. Slightly modifying the formula introduced in Aoyama et al. (2008), we can compute a_is by

$$a_i = \frac{1}{n}\binom{n}{i\,,\; n-2i+1\,,\; i-1}, \quad i = 1,2,...,\lceil (k_l-1)/2 \rceil\,. \tag{5.2.8}$$

According to Table 5.1, the total number of paths, all of which are observed as

$$\boldsymbol{x}_{obs} = (0,0,0,0,0,0,0,0,0,0,0,1),$$

is $\sum a_i = 5798$. Assuming $h = 0.001$ seconds, this example shows that even for a small period of 12 milliseconds (12 steps), the number of possible paths (unobservable) can be quite large. Since the step size is the infinitesimal value $h > 0$, the number of steps is usually a large number in real data. Furthermore, the numbers α_i, β_i and γ_i get larger and larger as the interspike interval increases. Since the parameters p_-, p_0 and p_+ are probabilities, raising them to large powers introduces round-off error to the estimation problem, which will be discussed later. Moreover, Equation (5.2.8) is computationally intensive for large values of n, and it is not recommended to compute the coefficients a_i in such cases.

The derivation of the likelihood function in Equation (5.2.7) is based on the assumption that the beginning of the observation window is at the first spike. Otherwise, the first path may not start at zero since the state of the process is not known at the beginning of the observation window. Therefore, an extra term, which calculates the likelihood of all the possible paths from the start point to the first spike should be imposed on the likelihood function, Equation (5.2.7). Similar adjustment should be applied to the interval between the last spike and the end of the observation window. We will address this subtle point and take it into consideration in the computationally efficient method introduced next for the calculation of the loglikelihood function.

5.2.1 A computationally tractable method for parameter estimation

The round-off error in the calculation of the likelihood function using Equation (5.2.7) increases significantly as the duration of the inactivity periods, i.e. the number of consecutive zeros in the observed data increases. This is because the longer the inactive period, the more possible paths, and the larger the powers (α, β, γ) of probabilities (p_0, p_+, p_-) in Equation (5.2.7). Furthermore, calculating the number of possible paths a_i, gets computationally intensive for relatively long periods with no spikes. To have a more accurate and tractable parameter estimation procedure, we will introduce a recursive algorithm. It is both computationally efficient and easy to implement.

Discretizing time, let $\{Z_i = \sum_{j<i}(X_j - X_j^{obs}) + X_i^{obs},\ i = 0, 1, ...\}$ represent the trajectory of the Skellam Process with Resetting (SPR). An example of this sample path was plotted in Figure 4.3. Let m be the current state of the SPR, $m \in \{1, 0, -1, -2, ...\}$, and let k be the number of steps from current time to the next time that $Z_i = 1$ i.e.,

$$k := \min\left\{j :\ X_j^{obs} = Z_j = 1,\ j > \text{current time}\,(i)\right\}$$

where $k \in \{1, 2, 3, ...\}$. The quantity

$$\begin{aligned} P(k|m) &= \Pr\Big(Z_k = 1 \Big| Z_i = m\Big) \\ &= \Pr\Big(X_k^{obs} = 1,\ X_j^{obs} = 0 \text{ for } m < j < k \Big| Z_i = m\Big) \end{aligned}$$

denotes the probability of observing the next spike in exactly $k > 0$ steps while the process is currently at $m \leq 1$. Conditioning on the first step, and using the independence assumption among unobservable X_is we have

$$P(k|m) = \begin{cases} p_- P(k-1|m-1) + p_0 P(k-1|m) + p_+ P(k-1|m+1) & \text{if } \ m < 0 \\ p_- P(k-1|-1) + p_0 P(k-1|0) & \text{if } \ m = 0 \end{cases} \tag{5.2.9}$$

with the initial conditions

$$P(k|m) = \begin{cases} p_+ & \text{if } \ k = 1 \ \& \ m = 0 \\ 0 & \text{if } \ k = m = 1 \ \text{ or } \ m \leq -k \\ P(k-1|0) & \text{if } \ m = 1 \end{cases}$$

The last term in the initial conditions above is due to the resetting where process is set to zero, accounting for neural refractoriness after each and every spike, Equation (4.3.8). Now, let us write the likelihood function of the parameters based on the

recursive function P introduced in Equation (5.2.9). After the first spike, the process is reset and starts at $m = 0$. Based on a fixed step-size h, let k_1 show the number of steps between the first and the second spikes. The probability of observing the second spike in exactly k_1 steps from the first spike is $P(k_1|1)$. However, for any given parameter values Equation (5.2.6) also calculates the probability of this event. Therefore

$$P(k_1|1) = \sum_{i=1}^{\lceil (k_1-1)/2 \rceil} a_i \; p_-^{\alpha_i} \; p_0^{\beta_i} \; p_+^{\gamma_i}.$$

Now suppose that k_2 steps exist between the second and the third spikes, k_3 steps between the third and the fourth spikes and so on. If there are n spikes in the spike train, the conditional independence of the spikes implies that

$$\begin{aligned} \prod_{i=1}^{n-1} P(k_i|1) &= \prod_{l=1}^{n-1} \sum_{i=1}^{\lceil (k_l-1)/2 \rceil} a_i \; p_-^{\alpha_i} \; p_0^{\beta_i} \; p_+^{\gamma_i} \\ &= L\Big(p_-, p_+; \boldsymbol{x}_1^{obs}, ..., \boldsymbol{x}_n^{obs}\Big), \end{aligned}$$

or equivalently the loglikelihood function based on the observation period from the first spike to the last spike is

$$\ell\Big(p_-, p_+; \boldsymbol{x}_1^{obs}, ..., \boldsymbol{x}_n^{obs}\Big) = \sum_{i=1}^{n-1} \log\Big(P(k_i|1)\Big) \tag{5.2.10}$$

5.2.2 Two intervals: before the first and after the last spike

Notice that in the loglikelihood (5.2.10) the two time intervals before the first spike and after the last spike are ignored. To take these periods into account, two terms should be added to Equation (5.2.10). Let k_0 shows the number of steps before the first spike, i.e.

$$k_0 := \min\Big\{j : \; X_j^{obs} = Z_j = 1, \; j > 0\Big\}.$$

Conditioning on the initial value of the process we have

$$\begin{aligned} &\Pr\Big(X_{k_0}^{obs} = 1 \; , \; X_j^{obs} = 0 \;\; \text{for } j < k_0\Big) \\ &\qquad = \sum_m \Pr\Big(X_{k_0}^{obs} = 1, \; X_j^{obs} = 0 \text{ for } j < k_0 \Big| Z_0 = m\Big) \Pr\Big(Z_0 = m\Big) \\ &\qquad = \sum_m P(k_0|m) \Pr\Big(Z_0 = m\Big) \end{aligned}$$

Since the first spike has happened in k_0 steps, $Z_0 \in \{1, 0, -1, -2, ..., -k_0 + 1\}$. Notice that the probability function $\Pr(Z_0 = m)$ is unknown and needs to be approximated.

We have tried utilizing the uniform distribution, as well as the stationary distribution of the chain, Equation (4.4.11), for $\Pr(Z_0 = m)$. It turns out that the stationary distribution outperforms the uniform distribution in terms of the fit. Having said that, note that stationary distribution is not necessarily the best approximation as we don't know if the chain is at equilibrium. Using the stationary distribution of the chain, Equation (4.4.11)], we have

$$\Pr\Big(X_{k_0}^{obs} = 1\,,\, X_j^{obs} = 0 \text{ for } j < k_0\Big) = \sum_{m=-k_0+1}^{1} P(k_0|m)\,\pi_{(2-m)} \tag{5.2.11}$$

Now, let k_n show the number the steps after the last spike to the end of the observation interval. This time, the value of the process, m, at the end of the observation interval is unknown. However, a spike has just occurred, i.e. $Z_{k^*} = 1$ where $k^* = \sum_{i=0}^{n-1} k_i$, thus

$$\Pr\Big(X_{1+k^*}^{obs} = 0\,,\, X_{2+k^*}^{obs} = 0,\, ..., X_{k_n+k^*}^{obs} = 0 | Z_{k^*} = 1\Big) \;\; = \;\; 1 - \sum_{j=1}^{k_n} P(j|1) \tag{5.2.12}$$

Incorporating equations (5.2.11) and (5.2.12) in the loglikelihood (5.2.10) we get

$$\begin{aligned} \ell\Big(p_-, p_+; \boldsymbol{x}_1^{obs}, ..., \boldsymbol{x}_n^{obs}\Big) \;\; = \;\; & \log\left(\sum_{m=-k_0+1}^{1} P(k_0|m)\,\pi_m^*\right) \\ & + \sum_{i=1}^{n-1} \log\Big(P(k_i|1)\Big) + \log\left(1 - \sum_{j=1}^{k_n} P(j|1)\right) \end{aligned} \tag{5.2.13}$$

where $\lambda_1 h = p_+ + o(h)$ and $\lambda_2 h = p_- + o(h)$. Since $P(n|m)$ is a recursive function, one can compute the loglikelihood (5.2.13) by only one call to the function P if the intermediate steps which lead to calculation of $P(n|m)$ are saved. The following algorithm summarizes the loglikelihood calculations and optimization to get the ML estimates.

- **Step1:** Set the step size $h = h_0$ ($\leq$ refractory period),
- **Step2:** For a spike train of n spikes, calculate the discretized interspike intervals $k_1, k_2, ..., k_{n-1}$, which are the number of steps between consecutive spikes, as well as the number of steps before the first spike, k_0, and the number of steps after the last spike k_n,
- **Step3:** Set $k_{max} = \max(k_i,\ i = 0, 2, ..., n)$. Using equations (5.2.9) and (5.2.10), define the matrix B such that $B_{i,j} = P(j| - i + 2)$, $i, j = 1, 2, ..., n + 1$. The rows and the columns of B represent, respectively, the values of m and n. As an example $P(5|-2) = B_{45}$.
- **Step4:** Optimize the bivariate loglikelihood (5.2.13), whose values will be computed based the elements of the matrix B.

If $\lambda_2 = 0$, the Skellam model is reduced to the conventional Poisson model with the effect of refractoriness (resetting). Not only does Skellam process allow for both negative and positive correlation coefficients among spike trains, it is also an insightful model in terms of neural inhibition. The parameter λ_2 is a measure of integrated inhibitory postsynaptic potential. We will address the dependency structure of the Skellam model later in Chapter 6, where we extend the model to the multivariate case.

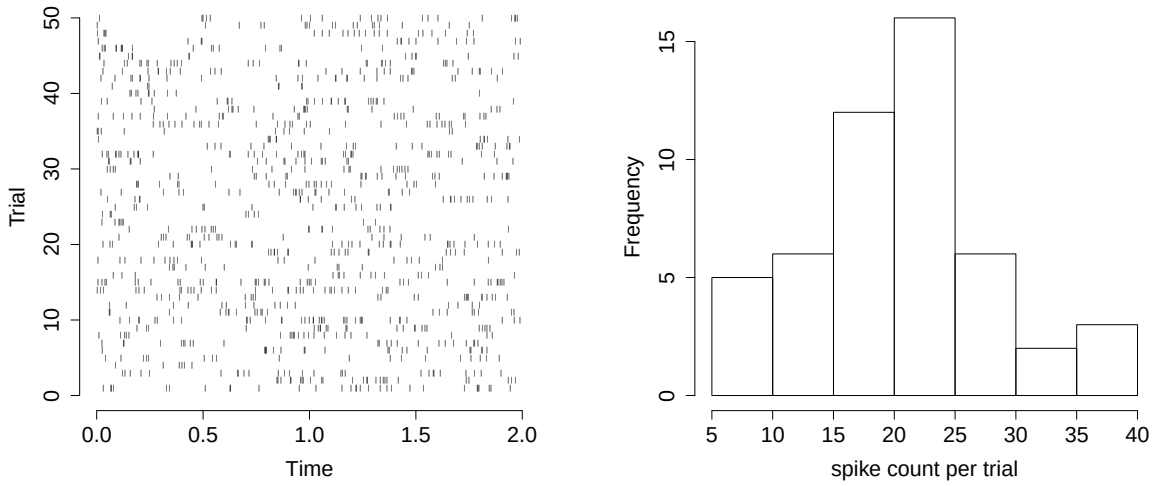
5.3 Data analysis

This section includes simulation studies as well as the analysis of the retinogeniculate synapse data (RGC and LGN neurons). The details of the real dataset and the experiment procedures were previously discussed in Chapter 3.

5.3.1 Simulation study

To check the performance of the Skellam model, we have simulated 50 spike trains from a SPR with parameters $(\lambda_1, \lambda_2) = (25, 15)$. Figure 5.2 shows the raster plot as well as the histogram of the spike count per trial.

We have employed the algorithm introduced above to estimate the parameters. The results of the analysis have been plotted in Figure 5.3.



(a) (b)

Figure 5.2: *Simulated data from a homogeneous SPR with parameters* $(\lambda_1, \lambda_2) = (25, 10)$. *panel (a) shows the raster plot of the data and panel (b) plots the histogram of spike count per trial.*

Based on $h = 0.001$, these estimates are $(\widehat{\lambda}_1, \widehat{\lambda}_2) = (25.19, 14.43)$ for the parameter values $(\lambda_1, \lambda_2) = (25, 15)$. To estimate the variability around the estimates, we have used 1000 nonparametric bootstrap samples each of size 50. Also provided, is the 95% confidence intervals based on the central limit theorem. Since the conditions of the central limit theorem seem to be reasonable here, the results based on the bootstrap method and C.L.T. are relatively similar. Table 5.2 summarizes the results.

Table 5.2: *Parameter estimates for the simulated data along with their 95% confidence intervals.*

Parameter	**Estimate**	**95% Confidence Interval**	
		(Bootstrap)	**(C.L.T.)**
$\lambda_1 = 25$	25.19	(22.61 , 28.34)	(22.29 , 28.09)
$\lambda_2 = 15$	14.43	(11.91 , 17.41)	(11.59 , 17.28)

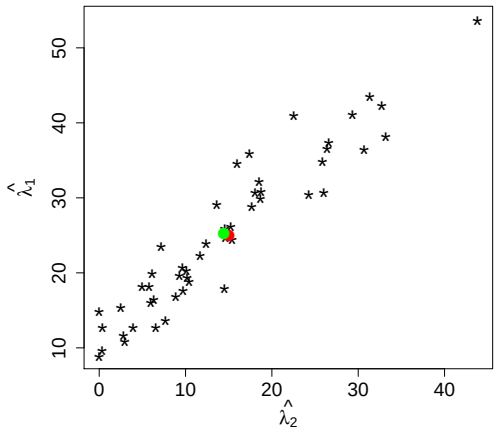


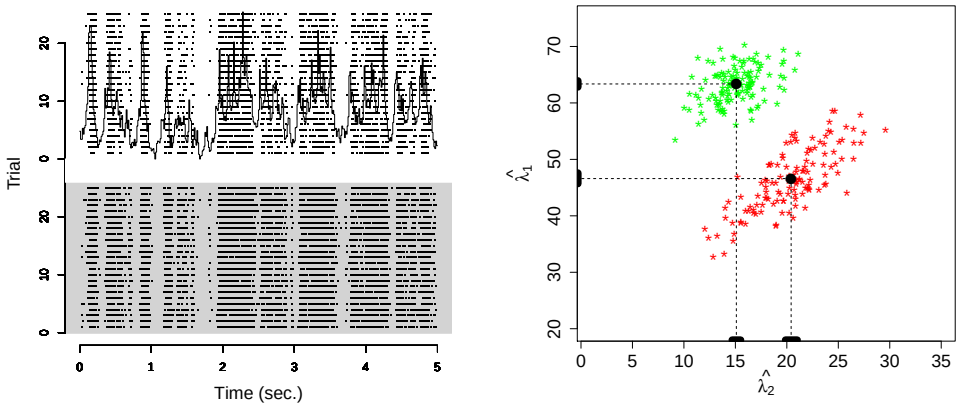
Figure 5.3: *Scatter plot of the parameter estimates* $\widehat{\lambda}_1$ *and* $\widehat{\lambda}_2$*. Each black asterisk shows the estimated parameters based on one of the 50 simulated trials. While the red dot shows the correct parameter value from which the 50 trials, i.e.* $(\lambda_1, \lambda_2) = (25, 10)$*, the green dot represents the average of the individual trials' estimates where* $(\overline{\widehat{\lambda}_1}, \overline{\widehat{\lambda}_2}) = (25.19, 14.43)$.

5.3.2 Real data analysis

We employ the similar dataset (retinogeniculate synapse) discussed in Chapter 3, which consists of 129 trials from an retinal ganglion cell RGC) as well as its paired LGN neuron. The estimated parameters based on each trial have been plotted in Figure 5.4.

Figure 5.4 clearly clusters the two neurons, with the RGC estimates sitting on top left of the LGN's. This shows a higher excitation $(\lambda_1 - \lambda_2)$ for the retinal ganglion cell (RGC) relative to its LGN connected neuron which is consistent with the higher spiking activity reported in Sincich et al. (2007) on this dataset. We generate 1000 nonparametric bootstrap samples each of size 129 trials to estimate the variability around the mean of the estimates. Also provided are 95% C.L.T. confidence intervals. Table 5.3 summarizes the findings.

It is known that not all the neural spikes in RGC get transmitted to the LGN, Sincich et al. (2007); Carandini et al. (2007). In particular, Carandini et al. (2007) showed that in the absence of any other inputs, it takes two Excitatory Postsynaptic Potentials (EPSP) occurring within 30 milliseconds to drive one LGN spike. This "selectivity" implies a lower spiking rate in LGN in each trial compared to the same trial for RGC.



(a) (b)

Figure 5.4: *Raster plot and estimated parameters based on the retinogeniculate synapse data. The green dots show the estimates from the RGC neuron, where their average is* $(\widehat{\lambda}_1, \widehat{\lambda}_2) = (63.36, 15.09)$. *The red dots represent those of the LGN cell data, where their average is* $(\widehat{\lambda}_1, \widehat{\lambda}_2) = (46.60, 20.42)$. *The black points show the average of the estimates and the thick bands on the axes represent marginal 95% confidence intervals.*

This biological property of paired RGC and LGN neurons has been reflected in statistical results where $\widehat{\lambda}_1^{(\mathrm{RGC})} > \widehat{\lambda}_1^{(\mathrm{LGN})}$ and $\widehat{\lambda}_2^{(\mathrm{RGC})} < \widehat{\lambda}_2^{(\mathrm{LGN})}$. Recall that increasing λ_1 and/or decreasing λ_2, increases the likelihood of spiking activity of a neuron.

5.4 Multiscale estimation

Since neural spike trains have biological characteristics from multiple time scales Ramezan et al. (2014); Nelson (2002); Kass and Ventura (2006), their analysis techniques should consider this multiscale nature of the neural spiking activity, which was addressed in details in Chapter 2. We employ the multiscale estimation algorithm of that chapter to estimate the time-varying functions $\lambda_1(t)$ and $\lambda_2(t)$. Notice that in this chapter we will only use the multiscale estimation algorithm without making any use of related theorems and/or upper bounds of the estimators' risk discussed in Kolaczyk and Nowak (2004). This is because the distribution of the number of spikes in the SPR framework does not have the so-called "cut" characteristic, hence the factorization of the likelihood function in the fashion of Chapter 2 is not possible in this setup. Refer to Kolaczyk and Nowak (2004); Barndorff-Nielsen (1976, 1978) for details on decoupling of the likelihood function and its underlying conditions. Compared to the Poisson pro-

Table 5.3: *Parameter estimates for the retinogeniculate data along with their 95% confidence intervals.*

Neuron	**Estimate**	**95% Confidence Interval**	
		(Bootstrap)	**(C.L.T.)**
RGC	$\widehat{\lambda}_1 = 63.36$	(62.82 , 63.92)	(62.80 , 63.92)
	$\widehat{\lambda}_2 = 15.09$	(14.69 , 15.51)	(14.70 , 15.49)
LGN	$\widehat{\lambda}_1 = 46.60$	(45.70 , 47.59)	(45.64 , 47.55)
	$\widehat{\lambda}_2 = 20.42$	(19.87 , 21.01)	(19.81 , 21.03)

cess, parameter estimation in SPR is computationally more intensive. For this reason we have employed recursive dyadic partitioning (RDP) as opposed to general recursive partitioning.

5.4.1 Tuning the parameters via cross-validation

The two parameters N, the number of intervals at the bottom of the recursive tree, and λ, the penalty factor, are the tuning parameters of the multiscale model discussed in Ramezan et al. (2014). We use similar cross-validation criteria to that of Chapter 2 which is, essentially, minimizing the integrated squared error loss. Let

$$f(t) = \frac{\lambda_1(t)}{\int_0^T \lambda_1(t)} \ , \quad g(t) = \frac{\lambda_2(t)}{\int_0^T \lambda_2(t)}.$$

Assume, also, that $\widehat{f}_i(t)$ and $\widehat{g}_i(t)$ are, respectively, the estimates of the functions $f(t)$ and $g(t)$, based on data from the i^{th} trial $i = 1, 2, ..., m$. Define

$$\widehat{f}^{-i}(t) = \sum_{j \neq i} \frac{\widehat{f}_j(t)}{m-1} \ , \qquad \widehat{g}^{-i}(t) = \sum_{j \neq i} \frac{\widehat{g}_j(t)}{m-1},$$
$$\widehat{f}(t) = \sum_{i=1}^{m} \frac{\widehat{f}_i(t)}{m} \ , \qquad \widehat{g}(t) = \sum_{i=1}^{m} \frac{\widehat{g}_i(t)}{m}.$$

The objective is to minimize the integrated square error losses $\int_0^T \Big(\widehat{f}(t) - f(t)\Big)^2 \, dt$ and $\int_0^T \Big(\widehat{g}(t) - g(t)\Big)^2 \, dt$. Based on these two criteria, we have

$$\begin{aligned} CV_{_1}(N, \lambda) &= \int_0^T [\widehat{f}(t)]^2 \, dt - 2\widehat{E}\Big(\widehat{f}(t)\Big) \\ &= \int_0^T [\widehat{f}(t)]^2 \, dt - \frac{2}{m} \sum_{i=1}^{m} \frac{\sum_{\ell=1}^{n_i} \widehat{f}^{-i}(t_{i\ell})}{n_i} \end{aligned} \tag{5.4.14}$$

$$
\begin{aligned}
CV_2(N,\lambda) &= \int_0^T [\widehat{g}(t)]^2\,dt - 2\widehat{E}\Big(\widehat{g}(t)\Big) \\
&= \int_0^T [\widehat{g}(t)]^2\,dt - \frac{2}{m}\sum_{i=1}^{m}\frac{\sum_{\ell=1}^{n_i}\widehat{g}^{-i}(t_{i\ell})}{n_i}
\end{aligned}
\tag{5.4.15}
$$

It is clear that the cross-validation is being done marginally on the two functions $\lambda_1(t)$ and $\lambda_2(t)$, therefore, the pair (N^*,λ^*) which minimizes $CV_1(N,\lambda)$ may not be the same as that of $CV_2(N,\lambda)$. To the best of our knowledge, no research study claims that the time-scale of the spiking activity of the inhibitory and excitatory presynaptic neurons are the same. Performing the cross-validation marginally, allows for having different structures/time-scales in the two functions $\lambda_1(t)$ and $\lambda_2(t)$. As a special case, constant $\lambda_2(t)$ and varying $\lambda_1(t)$ can be interpreted as relatively constant inhibitory activity, but variable excitatory activity.

5.4.2 Simulation study

We perform two simulation studies to see if the multiscale model fits the data well. First, we use the simulated data from Section 5.3.1 (homogeneous Skellam) to check if the multiscale algorithm merges all the bins and if it provides time-homogeneous estimates of the parameters λ_1 and λ_2. The tuning parameters λ (penalty factor) and N (number of the sub-intervals at the bottom of the dyadic tree), are chosen through the leave-one-trial-out cross-validation analysis described above. Cross-validation for SPR is computationally more intensive compared to the Poisson model. We have searched for the optimized pair over

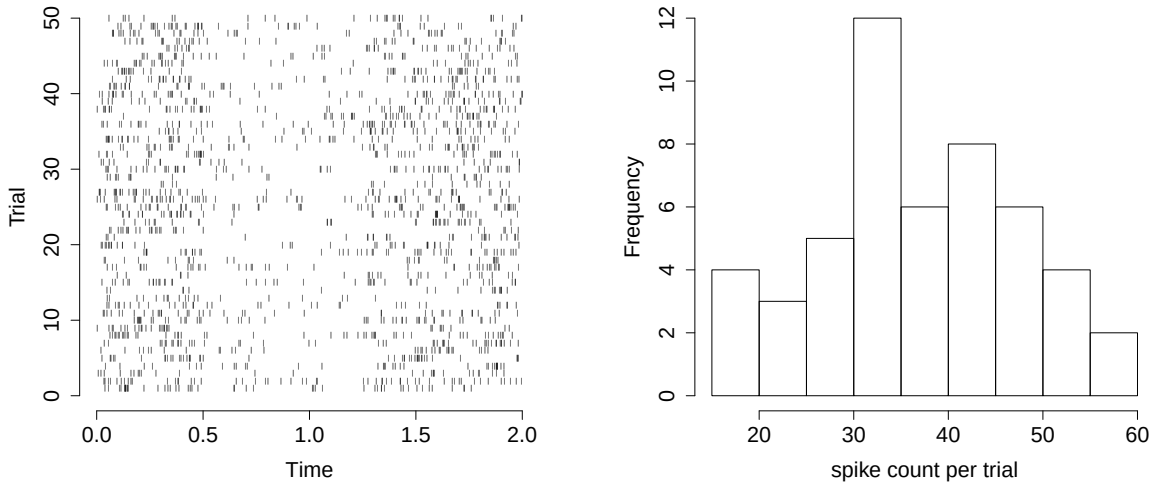
$$
\begin{aligned}
N &\in \{2^0, 2^1, ..., 2^7\}, \\
\lambda &\in \{0.005, 0.01, 0.015, 0.02, 0.03, 0.05, 0.07, 0.09, 0.1, 0.2, 0.3, ..., 1.5\}.
\end{aligned}
$$

The optimal values for λ and N are $(N^*,\lambda^*) = (2^2, 0.5)$, where the penalty factor $\lambda^* = 0.5$ is large enough to merge all sub-intervals to estimate constant values for both functions $\lambda_1(t)$ and $\lambda_2(t)$ across the simulation window $(0,2]$. As expected, the results of this study are the same as those presented in the time homogeneous case in Table 5.2. This shows that the multiscale estimator has correctly identified the time homogeneity of the functions $\lambda_1(t)$ and $\lambda_2(t)$.

In the second simulation, we have simulated an SPR with the following intensities,

$$
\lambda_1(t) = \begin{cases} 50 & 0 < t \le 0.5 \\ 10 & 0.5 < t \le 1.25 \\ 40 & 1.25 < t \le 2 \end{cases}
\qquad
\lambda_2(t) = \begin{cases} 20 & 0 < t \le 0.5 \\ 3 & 0.5 < t \le 1.25 \\ 15 & 1.25 < t \le 2 \end{cases}
\tag{5.4.16}
$$

Figure 5.5 shows the simulated data. Notice the clear sparsity in the raster plot during the middle intervals where both functions $\lambda_1(t)$ and $\lambda_2(t)$ have lower values. Applying the Cross-validation method of equations (5.4.14) and (5.4.15) the two "optimal" tuning parameters are $(N^*, \lambda^*) = (2^4, 0.05)$. Figure 5.6 plots the results of this analysis. Notice that the true functions $\lambda_1(t)$ and $\lambda_2(t)$ defined in Equation (5.4.16) both lie within their corresponding 95% confidence bands, as expected.

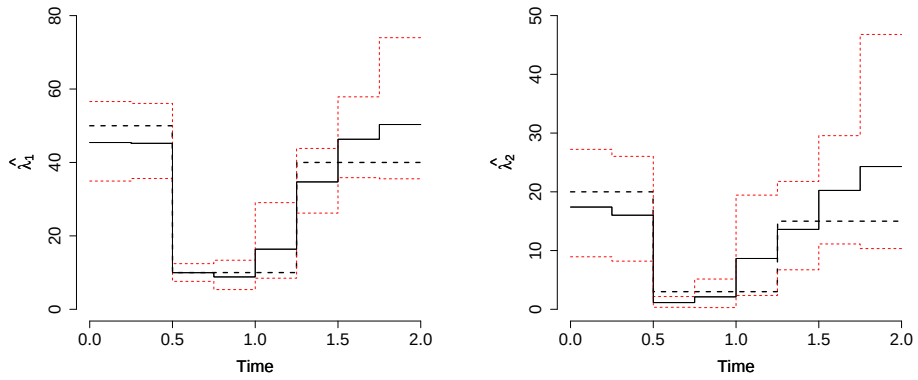


(a) (b)

Figure 5.5: *Simulated data from a inhomogeneous SPR with parameters functions* $\lambda_1(t)$ *and* $\lambda_2(t)$ *defined in* (5.4.16)*. panel (a) shows the raster plot of the data and panel (b) plots the histogram of spike count per trial.*

5.4.3 Real data analysis

Similar to Section 5.4.2, we have performed a multiscale analysis on the retinogeniculate synapse data. The optimal tuning parameters based on both LGN and RGC data are $(N^*, \lambda^*) = (2^6, 0.015)$. Figure 5.7 plots the average of the estimates $\widehat{\lambda}_1(t)$ and $\widehat{\lambda}_2(t)$ based on the 129 trials along with their 95% bootstrap confidence bands. Reading Figure 5.7, it is clear that the shape of the $\widehat{\lambda}_1(t)$ and $\widehat{\lambda}_2(t)$ are similar across the two neurons, which is expected from the similarity in the raster plots shown in Figure 5.4. The standard error of $\widehat{\lambda}_1(t)$ is smaller than that of $\widehat{\lambda}_2(t)$. This is due to the fact that more information about $\lambda_1(t)$ is present in the spike trains. Notice that $\lambda_2(t)$ is related to inhibition and negative values in SPR which are not observable in



(a) (b)

Figure 5.6: *The average multiscale estimates (solid black) based on the simulated data shown in Figure 5.5. The red dashed lines show the 95% bootstrap confidence bands. The black dashed lines show the true functions.*

spike trains. Figure 5.8 shows a simulated Skellam Process with Resetting with $\lambda_1(t)$ and $\lambda_2(t)$ plotted in Figure 5.7. The similarity between the raster plots of the real data and the simulated data shows that SPR model fits this data reasonably well.

In the SPR framework, $\lambda_1(t)-\lambda_2(t)$ is a measure of mean spiking rate, hence it contains some information about the stimulus signal to which the neuron has been exposed. In our estimation processes, stimulus signal was never taken into consideration. Comparing the estimate $\widehat{\lambda}_1(t) - \widehat{\lambda}_2(t)$ with the stimulus signal both plotted in Figure 5.9, we can see that most of the variability and structures in the stimulus signal have been captured by SPR model. This is a promising result in the performance of the SPR model. It is important to mention that we do not necessarily want to capture all the structures in the stimulus signal through the intensity function of the spike trains. The similarity in the shapes of the stimulus signal and the intensity function of neural spike trains is connected directly to the neural code, which is an unknown process. In short, the specific characteristics of this stimulus signal which are captured by the particular recorded neurons are not known.

The lower mean spiking rate in LGN relative to RGC (Figure 5.9b vs. Figure 5.9c) is consistent with the fact that the spiking activity in the RGC is about twice as much as its connected LGN neuron. Comparing the SPR with with the fit of inhomogeneous

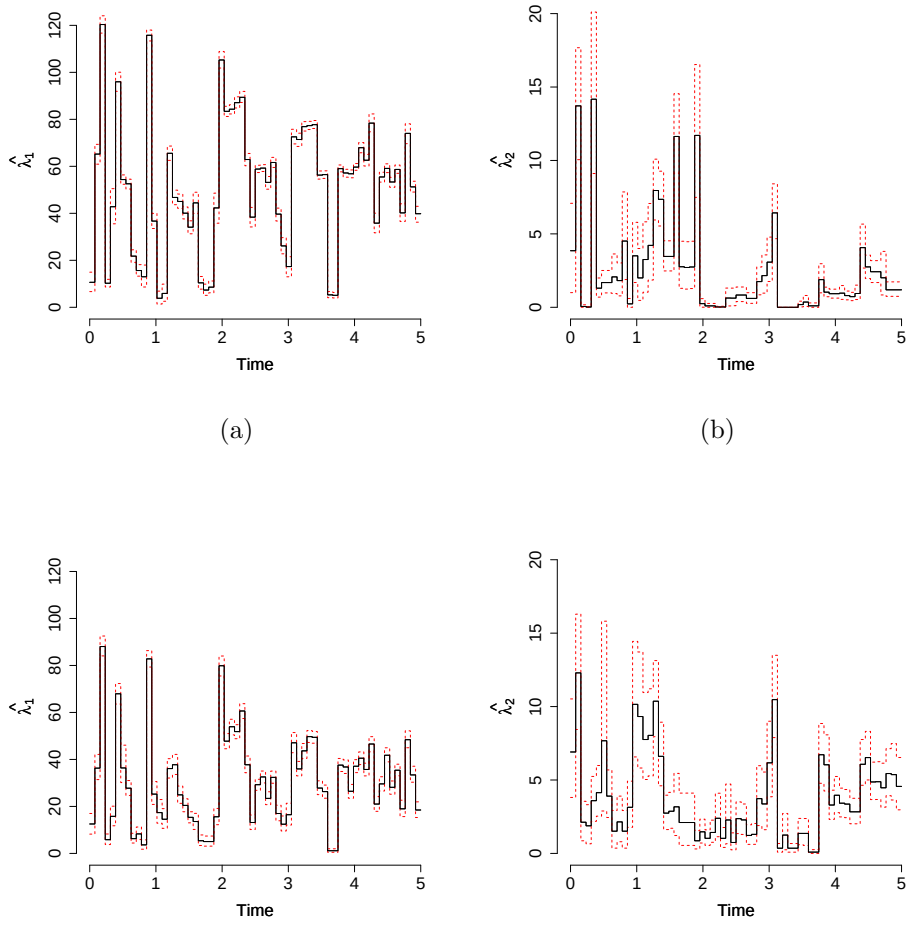
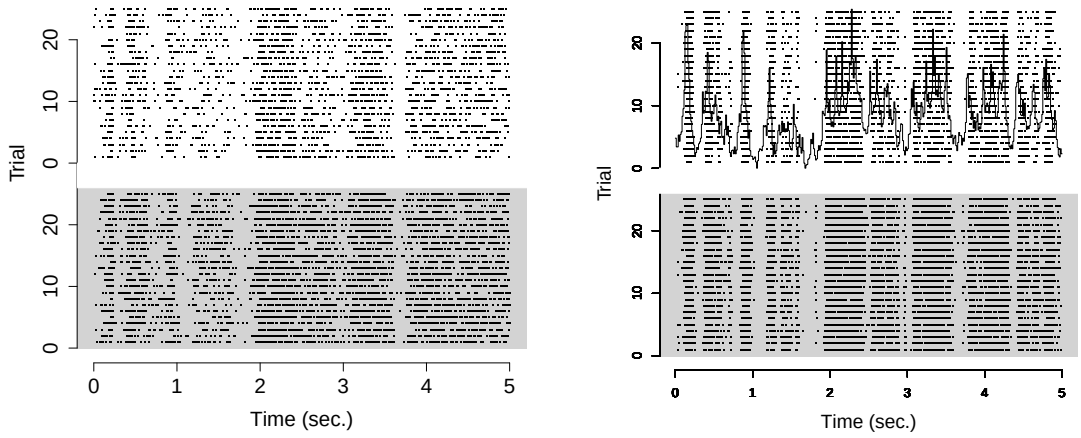


Figure 5.7: *The average multiscale estimate of the functions $\lambda_1(t)$ and $\lambda_2(t)$ based on the RG and LGN data. Panels (a) and (b) plot the estimates from the RGC data, while (c) and (d) show those of LGN. The red lines show the 95% bootstrap confidence bands.*

Poisson process, the results are quite comparable. However, there is less variability in the fit of SPR vs Poisson (see Figure 5.9).



(a) (b)

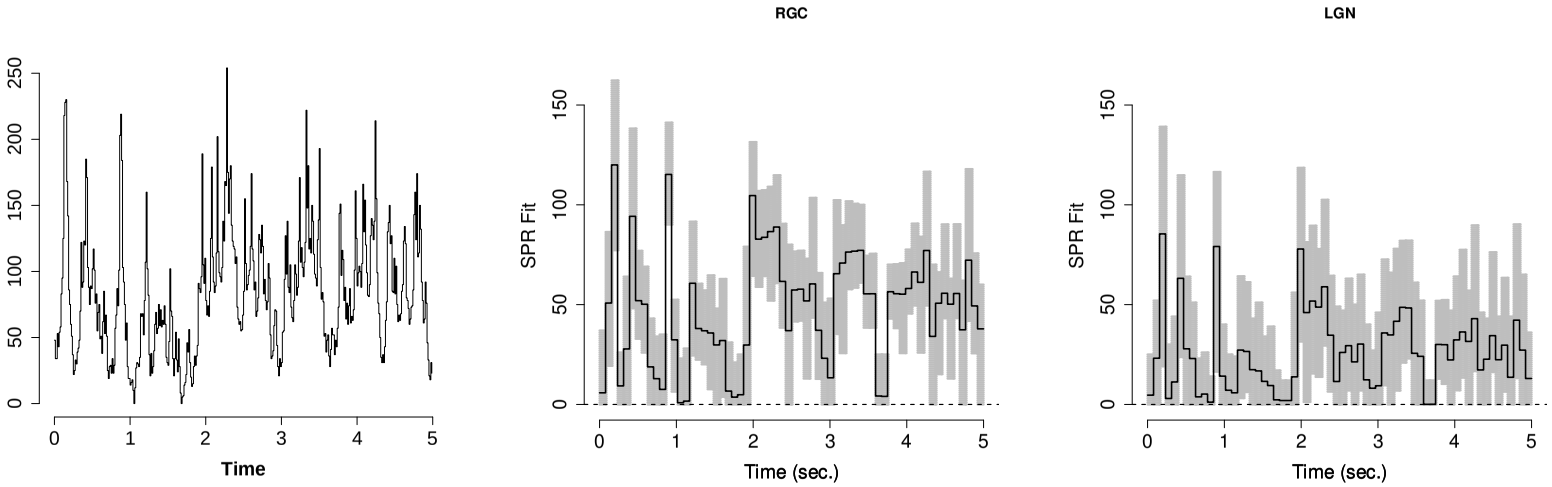
Figure 5.8: *Raster plots of 25 simulated SPRs with $\lambda_1(t)$ and $\lambda_2(t)$ plotted in Figure 5.7 as well as the 25 first trials from the real data. (a): simulated data and (b): real data. While the lower part (grey background) plots the RGC data, the white part shows LGN's. The curve in the upper part of panel (b) shows the scaled stimulus signal.*

5.5 Discussion

The derivation of the likelihood function and parameter estimation were discussed in this chapter. Based on the observable data (spike trains) we derived the likelihood function of the parameters/functions λ_1 and λ_2, where a computationally efficient algorithm for parameter estimation was developed. The multiscale estimation algorithm of Chapter 2 was tailored for the SPR framework. In a real data study we have shown that SPR is capable of reproducing the stimulus signal from the spiking activity of neurons.

Relative to the (inhomogeneous) Poisson process, SPR has one extra parameter/auxiliary function λ_2 making this new model both more flexible and biologically more insightful. Reading Figure 5.4b, λ_2 helps in clustering the two neurons. Furthermore, having the two quantities λ_1 and λ_2 makes the mean-variance relationship in the SPR more flexible relative to Poisson process.

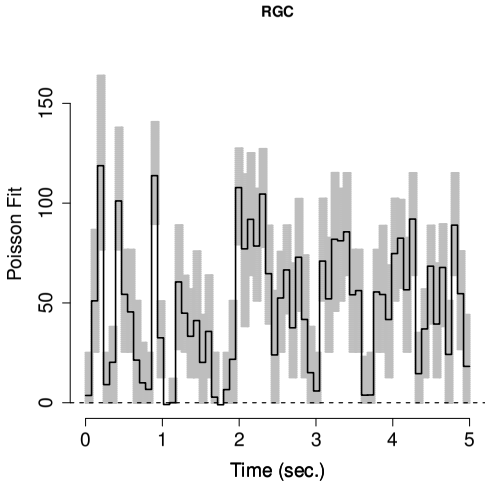
It is clear in Figure 5.7 that the confidence band around $\lambda_2(t)$ is wider than that of $\lambda_1(t)$. This is, of course, because the amount of available information about λ_1 is



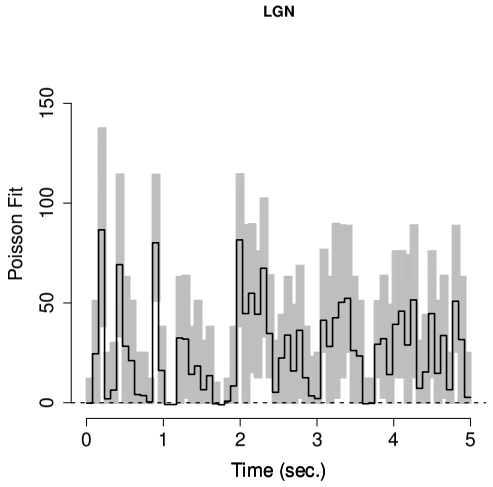
(a)

(b)

(c)



(d)



(e)

Figure 5.9: *The relationship between* $\lambda_1(t) - \lambda_2(t)$ *and the stimulus signal. The grey area shows the 95% quantile intervals based on the 129 repeated trials. (a): Stimulus signal to which the neurons have been exposed, (b):* $\widehat{\lambda}_1(t) - \widehat{\lambda}_2(t)$ *(SPR fit) for the RGC data, (c):* $\widehat{\lambda}_1(t) - \widehat{\lambda}_2(t)$ *(SPR fit) for the LGN data, (d): Poisson for the RGC data, (e): Poisson for the LGN data.*

significantly more than that of $\lambda_1(t)$. Notice that while $\lambda_1(t)$ is related to the neural excitation, hence the observable spikes, $\lambda_2(2)$ is related to the neural inhibition and the unobservable paths.

It is important to mention that the risk properties/boundaries discussed in Kolaczyk and Nowak (2004) are derived under Gaussian, Poisson and Multinomial models and those of the SPR model are yet to be investigated. Since the distribution of the number of spikes in SPR framework does not belong to the family of sum-symmetric power series distributions SSPSD family, the risk results of Kolaczyk and Nowak (2004) should be modified for this SPR framework.

Chapter 6

Multivariate Skellam Process With Resetting

6.1 Introduction

This chapter generalizes the results of Chapter 4 on the univariate Skellam process with resetting (SPR) to the multivariate case, where spike trains from multiple neurons can be simultaneously analyzed. We introduce the multivariate Skellam distribution and discuss some of its characteristics. Also introduced is the multivariate Skellam process in both homogeneous and inhomogeneous cases. We also generalize SPR to the *Multivariate Skellam Process with Resetting (MSPR)*. Parameter estimation in both simulation and real data analyses are discussed, and they provide promising results on the simultaneous analysis of multiple spike trains.

6.2 Conventional multivariate models

Studying the behaviour of a population of neurons and understanding their dependency structure is one of the main research streams in neuroscience, Rieke et al. (1997); Gerstner and Kistler (2002); Averbeck et al. (2006); Schneidman et al. (2006); Grun and Rotter (2010). However, simultaneous analysis of multiple neural spike trains is a challenging problem, Brown et al. (2004); Kass et al. (2005). In the case of single neurons, the estimate of the intensity function of spiking activity is a popular tool to perform inference about neural activity, Dimatteo et al. (2001); Kass and Ventura (2001); Behseta and Chenouri (2011); Ramezan et al. (2014). Histogram-based smoothing techniques such as Peristimulus Time Histogram (PSTH), Palm et al. (1988); Kass et al. (2003), BARS, Dimatteo et al. (2001), or kernel estimates, Park et al. (2012, 2013) are com-

mon techniques in modelling and analysis of neural spike trains. Poisson models are popular and widely used in the literature, see Chapter 2, but in order to perform simultaneous inference based on spike trains from a population of p neurons within a Poisson framework, one might use a multivariate distribution for the joint spike counts whose marginal distributions are Poisson. Such a distribution is called a *multivariate Poisson distribution*, (Johnson et al., 1997, p.139). Based on this definition, the multivariate Poisson probability distribution function is not uniquely defined. For different representations in the bivariate case refer to (Kocherlakota and Kocherlakota, 1992, p.100) and Lakshminarayana et al. (1999). According to Marshall and Olkin (1985), a bivariate Poisson distribution was first derived in McKendrick (1925) as the solution to a differential equation arising in biological applications. This distribution is, in fact, the joint distribution of the two random variables

$$T_1 = Y_1 + Y_2 \sim \text{Pois}(\lambda_1 + \lambda_2), \tag{6.2.1}$$
$$T_2 = Y_2 + Y_3 \sim \text{Pois}(\lambda_2 + \lambda_3), \tag{6.2.2}$$

where $Y_i \overset{\text{ind.}}{\sim} Poi(\lambda_i),\ i = 1, 2, 3$, Johnson et al. (1997); McKendrick (1925); Marshall and Olkin (1985); Kocherlakota and Kocherlakota (1992) and references therein for more details. The main problem with this approach is that since $\text{Cov}(T_1, T_2) = \lambda_2 > 0$, negative correlation is not allowed. The book of (Kocherlakota and Kocherlakota, 1992, pp.87-99) studied the properties of a bivariate Poisson distribution whose probability generating function is of the form

$$M(t_1, t_2) = \exp\Big\{\lambda_1(t_1 - 1) + \lambda_2(t_2 - 1) + \lambda_3(t_1 - 1)(t_2 - 1)\Big\}, \tag{6.2.3}$$

where they addressed the details of the limited dependency structure of this model. Expanding on the work of Griffiths et al. (1979), (Kocherlakota and Kocherlakota, 1992, p.100) introduced the probability generating function of a class of bivariate Poisson distributions, which does allow for some negative correlation at the cost of sacrificing infinite divisibility. Lakshminarayana et al. (1999) also introduced a bivariate Poisson distribution through the product of two marginal Poisson distributions and a multiplicative factor, which allows for negative correlations. However, the last two references have only addressed the bivariate case, and do not have clear generalizations to multivariate Poisson distribution.

Generalizing the bivariate model introduced in Equations (6.2.1) and (6.2.2), Karlis and Meligkotsidou (2005) introduced the multivariate Poisson random vector $(T_1, T_2, ..., T_k)$ where T_is are the sum of independent Poisson random variables. This generalization

does solve the problem of having a "multivariate Poisson distribution", but it does not allow for negative correlations. The book of Grun and Rotter (2010) addresses the correlations between spike trains in the same manner described in Equations (6.2.1) and (6.2.2). The positive correlation is too limiting in the analysis of neural spike trains. The biological concept of inhibitory and excitatory neurons, which implies the existence of negatively correlated spike trains, Ecker et al. (2010); Renart et al. (2010), is not properly formulated through the bivariate Poisson models introduced in Equations (6.2.1), (6.2.2), (6.2.3), and consequently the work of Karlis and Meligkotsidou (2005) on multivariate Poisson distribution.

The limitation in the dependency structure of the multivariate Poisson distribution motivates utilizing a different distribution for neural spike trains; however, it is convenient to develop a model which shares some properties with the Poisson model which has been extensively addressed in the literature. To accommodate this interest, we generalize the Skellam model of Chapter 4 to the multivariate case. We use a similar approach to that of Karlis and Meligkotsidou (2005) to introduce the Multivariate Skellam Distribution. We have recently noticed that a very special case of bivariate Skellam random variable has been discussed in Bulla et al. (2013).

6.3 Multivariate Skellam Distribution

Let $Y^{(\mathrm{I})} = (Y_1, ..., Y_p)^{\mathrm{T}}$ be a vector of p independent Skellam random variables, i.e.

$$Y_i \overset{\text{ind.}}{\sim} \mathrm{Sk}(\lambda_{i_1}, \lambda_{i_2}), \quad i = 1, .., p.$$

Also, let

$$Y^{(\mathrm{C})} = (Y_{12}, ..., Y_{1p}, Y_{23}, ..., Y_{2p}, Y_{34}, ..., Y_{3p}, ..., Y_{p-1p})^{\mathrm{T}}$$

be a vector of $p(p-1)/2$ zero-mean independent Skellam random variables, i.e.,

$$Y_{ij} \overset{\text{ind.}}{\sim} \mathrm{Sk}(\gamma_{ij}, \gamma_{ij}) \quad i = 1, ..., p-1,\ j = 2, ..., p,\ i < j.$$

Furthermore, let us assume that the two vectors $Y^{(\mathrm{I})}$ and $Y^{(\mathrm{C})}$ are also independent of each other. We introduce the p-variate Skellam distribution through the vector

$$Y = \begin{pmatrix} Y^{(\mathrm{I})} \\ Y^{(\mathrm{C})} \end{pmatrix}$$

and the matrix $C = (I\ ,\ A)$, where I is the identity matrix of size p, and A is a $p \times \frac{p(p-1)}{2}$ matrix whose elements are -1, 0 or 1. We will see later that while the random variables

in the vector $Y^{(\mathrm{I})}$ capture the contribution of individual neurons to their firing rates (I stands for individual), the ones in $Y^{(\mathrm{C})}$ model the pairwise associations (C stands for combinations). The vector $Z = (Z_1, ..., Z_p)^{\mathrm{T}}$ defined by $Z = CY$ is called a *p-variate Skellam random vector* and follows a *p-variate Skellam distribution*. We show this by

$$Z \sim MSk_p\Big(\lambda_{ij},\, \gamma_{ij}\Big),\; i, j \in \{1, ..., p\},\; i < j \text{ in } \gamma_{ij}, \tag{6.3.4}$$

where the indexes i and j are not equal for the parameters γ_{ij}. This is because γ_{ij} is the parameter associated with the distribution of the random variable Y_{ij}, which, essentially, captures the correlation between Z_i and Z_j in the vector Z (see Equation (6.3.9)). We have

$$\begin{aligned} Z &= C\,Y \\ &= \big(I\,,\, A\big)\begin{pmatrix} Y^{(\mathrm{I})} \\ Y^{(\mathrm{C})} \end{pmatrix} \\ &= Y^{(\mathrm{I})} + AY^{(\mathrm{C})}. \end{aligned} \tag{6.3.5}$$

The mean and the covariance matrix of the multivariate Skellam random vector, which are respectively $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)^{\mathrm{T}}$ and $\boldsymbol{\Sigma}$ are

$$\begin{aligned} \boldsymbol{\mu} &= E\big(Y^{(\mathrm{I})}\big)\,, \\ \boldsymbol{\Sigma} &= \Sigma^{(I)} + A\Sigma^{(C)}A^{\mathrm{T}}\,, \end{aligned}$$

where $\Sigma^{(I)}$ and $\Sigma^{(C)}$ are the covariance matrices of $Y^{(\mathrm{I})}$ and $Y^{(\mathrm{C})}$, respectively. We have

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \dots & \rho_{1p}\sigma_1\sigma_p \\ & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \dots & \rho_{2p}\sigma_2\sigma_p \\ & & \ddots & & \vdots \\ & & & & \sigma_p^2 \end{pmatrix}.$$

In this notation

$$\mu_i = \lambda_{i1} - \lambda_{i2}\,, \tag{6.3.6}$$

$$\sigma_i^2 = \lambda_{i1} + \lambda_{i2} + 2\gamma^{+}_{-(ii)}\,, \tag{6.3.7}$$

$$\rho_{ij} = \pm\frac{2\gamma_{ij}}{\sqrt{\sigma_i^2\sigma_j^2}} \quad i, j = 1, 2, ..., p\,, \tag{6.3.8}$$

in which,

$$\gamma^{+}_{-(ii)} = \gamma_{1i} + \gamma_{2i} + \cdots + \gamma_{i-1i} + \gamma_{ii+1} + \gamma_{ii+2} + \cdots + \gamma_{ip}.$$

The random vector Z can be written as

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix} = \begin{pmatrix} Y_1 \pm Y_{12} \pm Y_{13} \pm Y_{14} \pm \cdots \pm Y_{1p} \\ Y_2 \pm Y_{12} \pm Y_{23} \pm Y_{24} \pm \cdots \pm Y_{2p} \\ Y_3 \pm Y_{13} \pm Y_{23} \pm Y_{34} \pm \cdots \pm Y_{3p} \\ \vdots \\ Y_p \pm Y_{1p} \pm Y_{2p} \pm Y_{3p} \pm \cdots \pm Y_{p-1\,p} \end{pmatrix}, \tag{6.3.9}$$

where the signs are determined by the elements of matrix A.

While $Y^{(\mathrm{I})}$ in Equation (6.3.5) includes independent random variables, each of which are specific to one and only one random variable Z_i, $i = 1, ..., p$, $Y^{(\mathrm{C})}$ includes random variables which determine the covariances between each pair (Z_i, Z_j). It is clear that $\Sigma^{(I)}$ is a diagonal matrix, i.e.

$$\Sigma^{(I)} = \begin{pmatrix} \lambda_{11} + \lambda_{12} & 0 & \dots & 0 \\ 0 & \lambda_{21} + \lambda_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{p1} + \lambda_{p2} \end{pmatrix},$$

thus, if the elements of A are all set to zero, the random variables in Z are uncorrelated. Since the random vector $Y^{(\mathrm{C})}$ has mean zero, we have $E(Z) = E\big(Y^{(\mathrm{I})}\big)$. This is implied by defining each of the Skellam random variables Y_{ij} through only one parameter γ_{ij}, which turns out to be necessary for identifiability of the parameters. The probability mass function of the random vector Z defined in Equation (6.3.5) is

$$\begin{aligned} p(z_1, \dots, z_p) &= \sum_{y_{12}} \cdots \sum_{y_{p-1\,p}} \Big[p(z_1, \dots, z_p \,|\, y_{12}, y_{13}, \dots, y_{p-1\,p}) \prod_{i,j} p_{\gamma_{ij},\gamma_{ij}}(y_{ij}) \Big] \\ &= \sum_{y_{12}} \cdots \sum_{y_{p-1\,p}} \Big[p_{\lambda_{11},\lambda_{12}}(z_1 \mp y_{12} \mp \cdots \mp y_{1p}) \times \\ &\qquad p_{\lambda_{21},\lambda_{22}}(z_2 \mp y_{12} \mp \cdots \mp y_{2p}) \times \cdots \times \\ &\qquad p_{\lambda_{p1},\lambda_{p2}}(z_p \mp y_{1p} \mp \cdots \mp y_{p-1\,p}) \times \prod_{i,j} p_{\gamma_{ij},\gamma_{ij}}(y_{ij}) \Big]. \end{aligned} \tag{6.3.10}$$

where $p_{\lambda_{i2}\lambda_{i2}}$ and $p_{\gamma_{ij}}$ are the probability mass functions of Skellam random variables $\mathrm{Sk}(\lambda_{i1}, \lambda_{i2})$ and $\mathrm{Sk}(\gamma_{ij}, \gamma_{ij})$, respectively.

As a special case, let us investigate the bivariate Skellam random vector with +,+ configuration, i.e.

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Y_1 + Y_{12} \\ Y_2 + Y_{12} \end{pmatrix} .$$

The elements of this vector are positively correlated. Following Equation (6.3.10) we have,

$$\begin{aligned}
\Pr\left[\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right] &= \Pr\Big(Y_1 + Y_{12} = z_1 \,,\, Y_2 + Y_{12} = z_2\Big) \\
&= \sum_{y_{12}=-\infty}^{+\infty} \Big\{ \Pr_{\lambda_{11}\lambda_{12}}\Big(Y_1 = z_1 - y_{12}\Big) \times \\
&\quad \Pr_{\lambda_{21}\lambda_{22}}\Big(Y_2 = z_2 - y_{12}\Big) \times \Pr_{\gamma_{12}}\Big(Y_{12} = y_{12}\Big) \Big\} \\
&= \exp\Big(-\lambda_{11} - \lambda_{12} - \lambda_{21} - \lambda_{22} - 2\gamma_{12}\Big) \times \\
&\quad \sum_{y_{12}=-\infty}^{+\infty} \Big\{ \sum_{x=0}^{\infty} \frac{\lambda_{11}^{z_1 - y_{12} + x} \lambda_{12}^{x}}{(z_1 - y_{12} + x)!x!} \times \\
&\quad \sum_{x=0}^{\infty} \frac{\lambda_{21}^{z_2 - y_{12} + x} \lambda_{22}^{x}}{(z_2 - y_{12} + x)!x!} \times \sum_{x=0}^{\infty} \frac{\gamma_{12}^{y_{12}+2x}}{(y_{12} + x)!x!} \Big\} .
\end{aligned}$$

For $w_1, w_2 \neq 0$, the joint probability generating function of *this* bivariate Skellam distribution $(+,+$ configuration) is

$$\begin{aligned}
G_Z(w_1, w_2) &= E\Big(w_1^{Z_1} w_2^{Z_2}\Big) \\
&= E\Big(w_1^{Y_1} w_2^{Y_2} (w_1 w_2)^{Y_{12}}\Big) \\
&= \exp\Big[-(\lambda_{11} + \lambda_{12} + \lambda_{21} + \lambda_{22} + 2\,\gamma_{12}) \\
&\quad + \lambda_{11} w_1 + \lambda_{21} w_2 + \gamma_{12} w_1 w_2 + \frac{\lambda_{12}}{w_1} + \frac{\lambda_{22}}{w_2} + \frac{\gamma_{12}}{w_1 w_2} \Big] .
\end{aligned} \tag{6.3.11}$$

The joint probability generating function for the negatively correlated pair $\binom{Y_1+Y_{12}}{Y_2-Y_{12}}$ is obtained by substituting $w_1 w_2$ by w_1/w_2 in Equation (6.3.11). Now consider the two vectors

$$
\begin{aligned}
Z^{(1)} &= \begin{pmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{pmatrix} = \begin{pmatrix} Y_1 + Y_{12} \\ Y_2 + Y_{12} \end{pmatrix} \\
Z^{(2)} &= \begin{pmatrix} Z_1^{(2)} \\ Z_2^{(2)} \end{pmatrix} = \begin{pmatrix} Y_1 + Y_{12} \\ Y_2 - Y_{12} \end{pmatrix}
\end{aligned}
$$

where

$$
Y_i \overset{\text{ind.}}{\sim} \text{Sk}(\lambda_{i1}, \lambda_{i2}) \perp Y_{12} \sim \text{Sk}(\gamma_{12}, \gamma_{12}).
$$

The symbol $\perp$ above shows independence. Following Equations (6.3.6), (6.3.7) and (6.3.8) the moment estimates of the five parameters $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12}$ based on either $Z^{(1)}$ or $Z^{(2)}$ are

$$
\begin{aligned}
\widehat{\lambda}_{11} &= \frac{1}{2}\Big(\widehat{\mu}_1 + \widehat{\sigma}_1^2 - |\widehat{\rho}_{12}|\widehat{\sigma}_1\widehat{\sigma}_2\Big), \qquad \widehat{\lambda}_{12} = \frac{1}{2}\Big(- \widehat{\mu}_1 + \widehat{\sigma}_1^2 - |\widehat{\rho}_{12}|\widehat{\sigma}_1\widehat{\sigma}_2\Big), \\
\widehat{\lambda}_{21} &= \frac{1}{2}\Big(\widehat{\mu}_2 + \widehat{\sigma}_2^2 - |\widehat{\rho}_{12}|\widehat{\sigma}_1\widehat{\sigma}_2\Big), \qquad \widehat{\lambda}_{22} = \frac{1}{2}\Big(- \widehat{\mu}_2 + \widehat{\sigma}_2^2 - |\widehat{\rho}_{12}|\widehat{\sigma}_1\widehat{\sigma}_2\Big), \\
\widehat{\gamma}_{12} &= \frac{1}{2}|\widehat{\rho}_{12}|\widehat{\sigma}_1\widehat{\sigma}_2. \qquad (6.3.12)
\end{aligned}
$$

Figure 6.1 shows the probability mass function of the two bivariate random vectors $Z^{(1)}$ and $Z^{(2)}$ where $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12}) = (7, 4, 6, 3, 15)$. The $+,-$ configuration in $Z^{(2)}$ ($+,+$ configuration in $Z^{(1)}$) implies negative (positive) correlation which is also clear in Figure 6.1. Notice that there exist four possible configurations two of which imply positive correlations.

6.4 Multivariate Skellam process

Extending the results of Chapter 4, this section introduces the multivariate Skellam process. For the sake of simplicity, we introduce the bivariate process with $+,+$ configuration (positive correlation between the marginal processes). The results hold true (with minor modifications) for the other configurations. The derivations for higher dimensions are similar. Clearly, as we get to the higher dimensions, the estimation problem becomes more and more computationally intense.

6.4.1 Bivariate homogeneous Skellam process

Similar to the univariate case, we define the *Bivariate Skellam process* as follows;

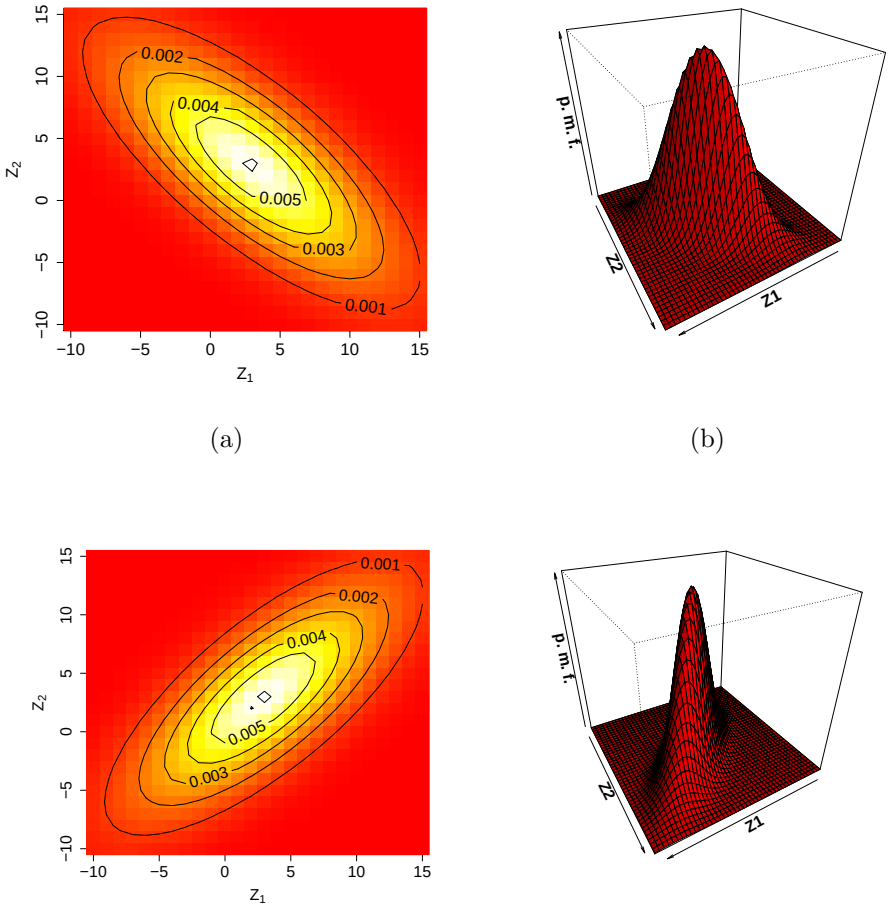


Figure 6.1: *The probability mass function of a bivariate Skellam random vector with parameters* $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12}) = (7, 4, 6, 3, 15)$. *The upper panel shows the* $+,-$ *configuration (negative correlation), while the lower panel illustrates the* $+,+$ *configuration (positive correlation).*

Definition 6.1 *An integer vector valued process* $\{MS_2(t) : t \geq 0\}$ *is said to be a* positively correlated homogeneous bivariate Skellam process *(*$+,+$ *configuration) with intensities* $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12})$, *all positive, if*

1. $MS_2(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2. *For all values* $t_1 < t_2$, *the random variable* $MS_2(t_2) - MS_2(t_1)$ *is independent of the times of incidents during* $[0, t_1]$ *(independent increments),*

3. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right] = 1 - (\lambda_{11} + \lambda_{12} + \lambda_{21} + \lambda_{22} + \gamma_{12})h + o(h),$

4. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right] = \lambda_{22}h + o(h),$

5. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] = \lambda_{21}h + o(h),$

6. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right] = \lambda_{12}h + o(h),$

7. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right] = \gamma_{12}h + o(h),$

8. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right] = o(h),$

9. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = \lambda_{11}h + o(h),$

10. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right] = o(h),$

11. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] = \gamma_{12}h + o(h).$

as $h \to 0$.

Alternatively, Bivariate Skellam process can be defined as follows.

Definition 6.2 *An integer vector valued process* $\{MS(t) : t \geq 0\}$ *is said to be a* homogeneous Bivariate Skellam Process *with parameters* $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12})$*, all positive, if*

1. $MS_2(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

2. $\{MS_2(t), t \geq 0\}$ *has independent increments,*

3. *For all values* $t, s \geq 0$, $MS(t+s) - MS(s) \sim MSk_2(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12})$.

Notice that the infinite divisibility of the bivariate Skellam distribution guarantees the existence of a stochastic process for the definitions above. This is because the bivariate Skellam process is also a Lèvy process.

Theorem 6.1 *Definitions 6.1 and 6.2 are equivalent.*

Proof: See Section 6.9.

We emphasize again that the results developed above are based on the $+,+$ configuration in the Skellam random variables, which results in positively correlated Skellam processes. The results based on other configurations are similar. Figure 6.2 shows how different configurations would affect the probability terms $\Pr\left[MS_2(t+h)-MS_2(t)=\binom{i}{j}\right]$ in Definition 6.1.

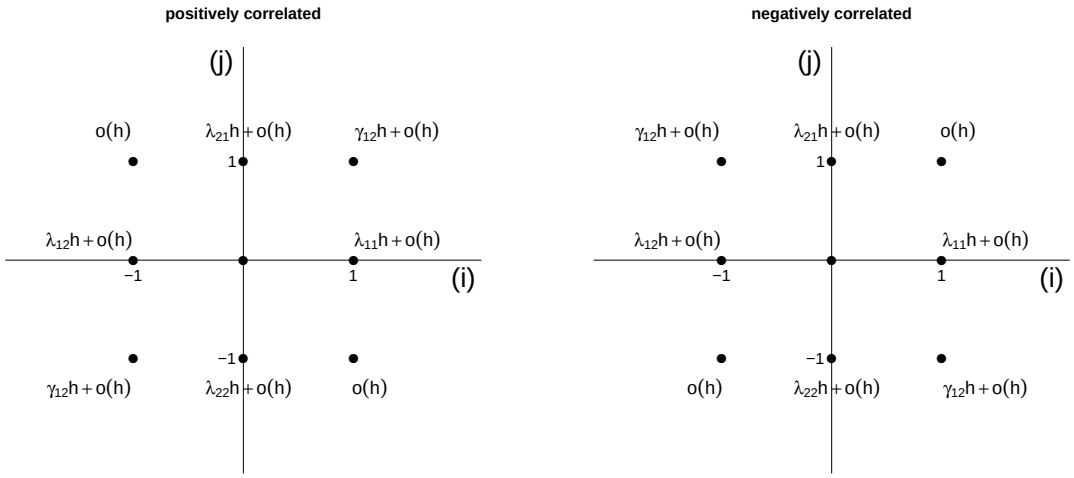


Figure 6.2: *The effect of correlation on probabilities in bivariate Skellam process*

It is noteworthy that Definition 6.2 does not depend on the configurations $(+,+)$, $(+,-)$, $(-,+)$ or $(-,-)$, however, to prove Definition 6.1 from this definition, $(+,+)$ configuration is assumed in the proof. Working with other configurations implies other versions of Definition 6.1. In particular, the configurations with positive correlation (+,+ and -,-) as well as the ones with negative correlation (+,- and -,+) are completely the same in terms of both the derivation and implementation.

6.4.2 Dimensionality of the parameter space

Based on the definition of the multivariate Skellam random vector Z [see Equation (6.3.9)] with entries

$$Z_j = Y_j \pm \underbrace{Y_{1j} \pm \cdots \pm Y_{j-1\,j}}_{j-1} \pm \underbrace{Y_{j\,j+1} \pm Y_{j\,j+2} \cdots \pm Y_{jp}}_{p-j},$$

each of the random variables $Z_j,\ j = 1, 2, ..., p$ is associated with $2 + (j-1) + (p-j) = p+1$ parameters. Notice that each pair $(Z_i, Z_j),\ i < j$ have the parameter γ_{ij} in common. Therefore, the total number of parameters of the p-variate Skellam distribution is

$$2 \times p + \big[(p-1) + (p-2) + \cdots + 1\big] = \frac{p(p+3)}{2}, \tag{6.4.13}$$

which is equal to the number of parameters of a p-variate normal distribution. Since the asymptotic distribution of the Skellam random variable is Gaussian (as the parameter values increase), this consistency in the number of parameters suggests that the multivariate Skellam distribution is not an over-parameterized or an under-parameterized formulation of the problem. For examples of Gaussian modelling of neural spike trains refer to Sompolinsky et al. (2001); Cunningham et al. (2008) and the references therein.

6.4.3 Inhomogeneous multivariate Skellam process

The inhomogeneous version of the multivariate Skellam process is defined as follows.

Definition 6.3 *An integer vector valued process* $\{MS_2(t) : t \geq 0\}$ *is said to be a* positively correlated inhomogeneous bivariate Skellam process *(+, + configuration) with non-negative time varying intensity functions* $\lambda_{11}(t)$, $\lambda_{12}(t)$, $\lambda_{21}(t)$, $\lambda_{22}(t)$ *and* $\gamma_{12}(t)$, *if*

1. $MS_2(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2. $\{MS_2(t), t \geq 0\}$ *has independent increments,*
3. $Pr\Big[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\Big] = 1 - \Big(\lambda_{11}(t) + \lambda_{12}(t) + \lambda_{21}(t) + \lambda_{22}(t) + \gamma_{12}(t)\Big)h + o(h),$

4. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right] = \lambda_{22}(t)h + o(h),$

5. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] = \lambda_{21}(t)h + o(h),$

6. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right] = \lambda_{12}(t)h + o(h),$

7. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right] = \gamma_{12}(t)h + o(h),$

8. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right] = o(h),$

9. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = \lambda_{11}(t)h + o(h),$

10. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right] = o(h),$

11. $Pr\left[MS_2(t+h) - MS_2(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] = \gamma_{12}(t)h + o(h).$

as $h \to 0$.

Alternatively, it can be defined as

Definition 6.4 *An integer vector valued process* $\{MS_2(t) : t \geq 0\}$ *is said to be a* inhomogeneous Bivariate Skellam Process *with non-negative time varying intensity functions* $\lambda_{11}(t)$, $\lambda_{12}(t)$, $\lambda_{21}(t)$, $\lambda_{22}(t)$ *and* $\gamma_{12}(t)$, *if*

1. $MS_2(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

2. $\{MS_2(t), t \geq 0\}$ *has independent increments,*

3. *For all values* $t, s \geq 0$, $MS_2(t+s) - MS_2(s)$ *follows a bivariate Skellam distribution with time varying intensity functions* $\int_s^{t+s} \lambda_{11}(y)\, dy$, $\int_s^{t+s} \lambda_{12}(y)\, dy$, $\int_s^{t+s} \lambda_{21}(y)\, dy$, $\int_s^{t+s} \lambda_{22}(y)\, dy$ *and* $\int_s^{t+s} \gamma_{12}(y)\, dy$.

Once again, the infinite divisibility of the increment probabilities (Skellam), or the fact that the process is a Lèvy process, guarantees the existence of a stochastic process defined above.
Similar to the homogeneous case we have,

Theorem 6.2 *Definitions 6.3 and 6.4 are equivalent.*

Proof: See Section 6.9.

6.5 Multivariate Skellam process with resetting (MSPR)

We introduced the Skellam process to model neural spike trains. In Chapters 4 and 5, univariate spike trains i.e., single neuron recordings, were modelled as realizations of the Skellam Process with Resetting (SPR), where we used the notion of records. A common sense generalization to the multivariate case (multiple neurons) is to model the multivariate spike trains as "records" of a multivariate Skellam process. However, multivariate records in this sense need to be properly defined. Although the literature on univariate records dates back to 1950s, Chandler (1952), the literature on multivariate records is not as rich. Most of the literature available on multivariate records belong to the past 20 years. Consider a sequence of independent random vectors $X_1, X_2, ...$ with common continuous distribution function F. The random variable X_n is called the *multiple maxima* if it is the componentwise sample maximum i.e., all components of $X_n - X_i$ are positive for $i < n$. Hashorva and Hüsler (2005) studied the asymptotic behaviour of this random variable. Substituting $X_n - X_i > 0$ for $i < n$ with $X_n - X_i \geq 0$ in this definition gives the *Pareto record* or *weak record*, Gnedin (2007); Hwang and Tsai (2010). Gnedin (2007) introduces yet another type of multivariate record called the *chain record*, which is also based on the partial ordering of the vectors $X_1, X_2, ...$ These three different types of multivariate records are studied in details in Hwang and Tsai (2010), where the mean and variance of record counts are derived and central limit theorems with convergence rates are established when the variance tends to infinity.

For the purpose of the joint analysis of spike trains from multiple neurons, we will, essentially, keep the records of the marginal processes in a vector, and treat this vector as the "multivariate record." As mentioned above, this is also referred to as the dominating record, Hwang and Tsai (2010), strong record, Gnedin (2007), or multiple maxima, Hashorva and Hüsler (2005). For the details of theoretical properties of this type of record refer to Hwang and Tsai (2010) and the references therein.

Definition 6.5 *Let* $\{S_0^{(i)}(t), t \geq 0\}$ $i = 1, ..., k$ *be* k *Skellam processes with records* $R_t^{(i)} = \max\{S_0^{(i)}(t^*),\ t^* < t\}$, *respectively. Also, assume that* $R_t^{(i)}$ *is independent of* $R_t^{(j)}$ *for* $i \neq j$. *Define* $MS_0(t)$ *and* MR_t *as follows,*

$$MS_0(t) = \begin{pmatrix} S_0^{(1)}(t) \\ \vdots \\ S_0^{(k)}(t) \end{pmatrix} \quad , \quad MR_t = \begin{pmatrix} R_t^{(1)} \\ \vdots \\ R_t^{(k)} \end{pmatrix}.$$

Multivariate Skellam Process with Resetting (MSPR) *is then defined to be*

$$MS = \Big\{ MS_0(t) - MR_t,\ t \geq 0 \Big\} \tag{6.5.14}$$

It is clear that the individual elements of the process defined in Equation (6.5.14) are univariate Skellam processes with resetting.

The assumption of independent marginal records in the definition above is, in fact, biologically plausible, because the absolute refractory period is intrinsic to the cell and is not affected by the synaptic input.

6.6 Parameter estimation

Although we generalized the recursive algorithm of Chapter 5 (see Section 6.9) to the bivariate case, we have noticed that it is too slow to be used in practice. This is because of the very large volume of the array analogous to the matrix M in the univariate version of the algorithm in Section 5.2.2. Even for the very small interspike interval of 1000 steps (1 second of one-millisecond steps), the array would have 10^9 elements, which makes this algorithm extremely slow. Moreover, the ultimate goal is to model more neurons than the bivariate case. Moment estimators or profile likelihood estimators are some alternatives for parameter estimation.

6.6.1 Method of moments and likelihood

It was shown in Section 4.5.3 that for large time scales, the number of spikes in an SPR model has a Skellam distribution. Therefore, we can use the method of moments to estimate the parameters. In the bivariate case, the set of Equations (6.3.12) can be used for parameter estimation. Not being domain-preservative (producing estimates outside the parameter space), the method of moments may produce unacceptable estimates for the parameters. Table 6.1 shows a few examples which confirm that the mapping from

the parameter space to the moment space is not onto. The bold negative values show that the moment estimates are not domain-preservative.

Table 6.1: *Mapping between the moment space and the parameter space for Skellam distribution.*

Moments					**Parameters**				
μ_1	σ_1^2	μ_2	σ_2^2	ρ	λ_{11}	λ_{12}	λ_{21}	λ_{22}	γ_{12}
50	150	30	90	-0.7	59.3	9.3	19.33	**-10.67**	40.67
60	103	37	176	0.5	47.84	**-12.16**	72.84	35.84	33.66
80	120	100	150	0.5	66.44	**-13.54**	91.46	**-8.54**	33.54

Notice that

$$\begin{aligned} \lambda_{12} &= \frac{1}{2}(\sigma_1^2 - \mu_1^2 - |\rho_{12}|\sigma_1\sigma_2)\,, \\ \lambda_{22} &= \frac{1}{2}(\sigma_2^2 - \mu_2^2 - |\rho_{12}|\sigma_1\sigma_2)\,, \end{aligned}$$

therefore, the unacceptable values occur when the over dispersion relative to Poisson model $(\sigma_i^2 - \mu_i)$ is dominated by the absolute value of the covariance between spike counts $(\rho_{12}\sigma_1\sigma_2)$. Our solution to this problem is to use a plug-in estimate of the parameter γ_{12} and to use the marginal likelihoods, which were derived in Chapter 5. The recursive function derived for marginal likelihood computations is (see Chapter 5)

$$P(k|m) = \begin{cases} p_- P(k-1|m-1) + p_0 P(k-1|m) + p_+ P(k-1|m+1) & \text{if } \ m < 0 \\ p_- P(k-1|-1) + p_0 P(k-1|0) & \text{if } \ m = 0 \end{cases}$$

with the initial conditions

$$P(k|m) = \begin{cases} p_+ & \text{if } \ k = 1 \ \& \ m = 0 \\ 0 & \text{if } \ k = m = 1 \ \text{ or } \ m \leq -k \\ P(k-1|0) & \text{if } \ m = 1 \end{cases}$$

in which, $p_+ = \lambda_{11}h$, $p_- = \lambda_{12}h$, and $p_0 = 1 - p_+ - p_-$ are the marginal probabilities associated with the 3-state random walk. In the bivariate case with $+,+$ configuration, these marginal probabilities are calculated by summing over the bivariate probabilities from Definition 6.1. In this case, we have

$$\begin{aligned} p_+ &= (\lambda_{11} + \gamma_{12})h + o(h)\,, \\ p_- &= (\lambda_{21} + \gamma_{12})h + o(h)\,, \\ p_0 &= 1 - p_+ - p_-\,. \end{aligned} \tag{6.6.15}$$

Now, we plug in the moment estimate of γ_{12} in the marginal likelihood function from Equation (6.1) with parameters p_+ and p_- defined in Equations (6.6.15) above, and maximize this "moment-profile likelihood function". We will use constraint optimization on this function because $p_+, p_- \geq \gamma_{12}h + o(h)$. The derivations/parameter estimation in the multivariate case (dimensions above 2) is similar.

6.7 Data analysis

We now employ the methods discussed above on both simulated data as well as the real retinogeniculate synapse data discussed in Chapters 3 and 5.

6.7.1 Simulation study

We have simulated 100 trials of a positively correlated bivariate Skellam process with resetting over the period $[0, T] = [0, 50]$ with parameters $(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \gamma_{12}) = (20, 10, 28, 8, 20)$ per unit time. Theoretically, the correlation coefficient among the spike counts of the two neurons over the above-mentioned period is 0.55, whose estimate from the data is 0.62 (S.E.=0.12). Based on the sample estimate of the correlation coefficient, the $+, +$ configuration is used. Table 6.2 summarizes the estimation results.

Table 6.2: *Simulation Results: parameter estimates for homogeneous Skellam process with resetting.*

Method	Parameter	Estimate (S.E.)
Moment-Profile Likelihood	$\lambda_{11} = 20$	16.12 (2.49)
	$\lambda_{12} = 10$	6.69 (2.45)
	$\lambda_{21} = 28$	22.35 (1.90)
	$\lambda_{22} = 8$	3.25 (1.96)
	$\gamma_{12} = 20$	19.37 (7.65)
Moments Method	$\lambda_{11} = 20$	21.35 (8.38)
	$\lambda_{12} = 10$	11.92 (8.34)
	$\lambda_{21} = 28$	26.44 (7.57)
	$\lambda_{22} = 8$	7.34 (7.54)
	$\gamma_{12} = 20$	19.37 (7.65)

From this simulation we can see that while the standard error of the moment-profile likelihood estimate is higher, method of moments provides better point estimates. Of

course, if the estimates from the moments method lie outside the parameter space, this method will not provide informative results. Based on the parameter values of the simulated bivariate neural spike trains, the mean and the variance of the spike counts per unit time (1 sec.) are $E(N_1) = 9.43$, $E(N_2) = 19.1$, $Var(N_1) = 72$ which are well-estimated by the estimates (true values lie within the 95% confidence intervals). Furthermore, the correlation coefficient between the two spike trains is 0.54 (based on the data), whose moment-profile likelihood estimate from the third column of Table 6.2 is 0.61 (S.E.=0.12). In summary, comparing the estimates to the true parameter values, we can see that the parameters of the multivariate Skellam model are estimated quite well.

6.7.2 Real data study

We apply the multivariate Skellam methodology on two neurons from the retinogeniculate synapse data. We fit the homogeneous Skellam model to the ON-centre Parvocellular LGN-RGC pairs, which consist of 129 repeated trials. According to Carandini et al. (2007), it takes two EPSPs within 30ms in RGC to drive one LGN spike. This shows that the two cells are positively correlated, so the +,+ configuration will be used. If such information was not available, we could easily test for the sign of the correlation coefficient between the two sets of spike trains. The raster plot of (the first 25 replications of) the data is shown in Figure 6.3 which confirms this positive correlation.

Table 6.3 summarizes the results of the method of moments and the profile likelihood model. Notice that the estimates from the moments method are not acceptable as they fall outside the parameter space.

The estimates of the mean and the variance of the spike counts across the 129 trials of the 5-second time window for the RGC data are 242.1 and 248.0, respectively. Similar estimates for the LGN data are 132.1 and 327.2. respectively. This shows that while the moment-profile likelihood estimate of the mean spike count is appropriate (236.67-0.075=236.6), it is over-estimating the variance for the RGC data ($236.67 + 0.075 + 2 \times 99.92 = 436.6$). This could be due to the strong inhomogeneity in the spike trains that we have analyzed, or because of a possible bias that the plug-in estimate of γ_{12} might have caused. Since the stimulus signal to which these neurons have been exposed varies significantly over time, fitting an inhomogeneous process seems more reasonable for this data.

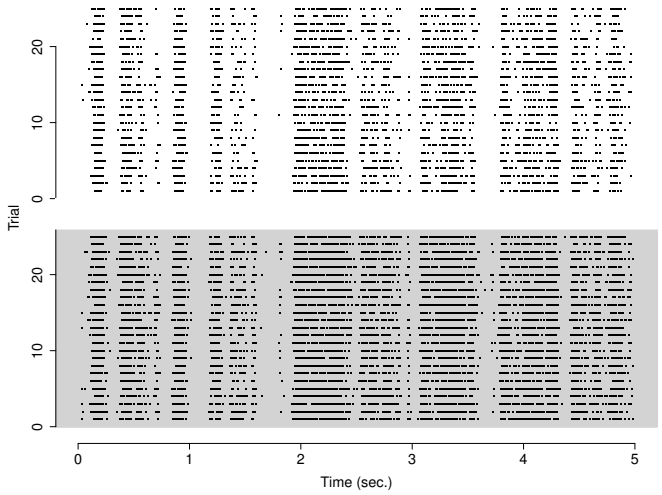


Figure 6.3: *The raster plot of the first 25 trials of the on-center parvocellular recordings. The grey panel represent the RGC recordings, while the white panels shows those of LGN.*

6.8 Discussion

This chapter generalized the univariate Skellam model to the multivariate case. The multivariate Skellam random vector and multivariate Skellam process with resetting (MSPR) were introduced. We proposed the so-called moment-profile likelihood method to estimate the parameters of the process. Our data analyses provided promising results on the performance of the multivariate Skellam model.

The general case of MSPR has $p(p+3)/2$ parameters, where p is the number of neurons. It can be shown that if the correlation coefficients among the elements of the random vector Z, Equation (6.3.9), is the same, i.e. $\rho_{ij} = \rho_0$ $i, j = 1, ..., p$, the "individual parameters" $(\lambda_{i1}, \lambda_{i2})$ $i = 1, ..., p$, and the "covariance parameters" γ_{ij} are functionally dependent. This reduces the model to a $2p$-dimensional parameter space as opposed to the original $p(p+3)/2$ dimensional one. It is noteworthy that while parameters γ_{ij} do not contribute to the mean spiking activity, they do contribute to the variance of the spike counts, allowing for over dispersion in the parallel analysis of neural spike trains.

Each element of Z in Equation (6.3.9) is expressed by a sum of independent terms Y_i and Y_{ij}, where these elements explain both the mean and the variability. These terms can be seen as the main effects and two-way interactions in an ANOVA set up. In

Table 6.3: *Parameter estimates for the RGC-LGN cells from 129 trials of an ON-centre Parvocellular pair. The unit time is assumed to be the 5-second time window.*

Method	Parameter	Estimate (S.E.)
Moment-Profile Likelihood	λ_{11}	236.67 (15.01)
	λ_{12}	0.075 (0.61)
	λ_{21}	138.23 (22.78)
	λ_{22}	8.14 (10.74)
	γ_{12}	99.92 (35.50)
Moments Method	λ_{11}	145.12 (25.0)
	λ_{12}	**-97.02** (25.33)
	λ_{21}	129.73 (27.83)
	λ_{22}	**-2.35** (27.23)
	γ_{12}	99.92 (39.62)

theory, more terms can be added to the model to count for the three-way up to m-way interactions in the model i.e., terms like Y_{ijk}, Y_{ijkl}, etc.

It was shown in Chapter 3 that the correlation between spike counts is a multiscale phenomena i.e., it depends on the bin-size during which the spikes are counted. It is important to mention that the time scale can also affect the moment estimates in our analysis above. In fact, one of the "next steps" in this area is to extend the multiscale algorithm of Chapters 3 and 5 so that it can be used in the multivariate case.

6.9 Proofs of the theorems and a parameter estimation algorithm

6.9.1 Proof of Theorem 6.1

First, we show that Definition 6.1 implies Definition 6.2. Let $\{MS_2(t), t \geq 0\}$ be a bivariate Skellam process. Therefore, $\{MS_2(t), t \geq 0\}$ satisfies the conditions listed in Definition 6.1. We show that for any values $s, t \geq 0$, the probability generating function of the random vector $MS_2(t+s) - MS_2(t)$ is the same as that of a bivariate Skellam random vector with intensities $\lambda_{11}t$, $\lambda_{12}t$, $\lambda_{21}t$, $\lambda_2 t$ and $\gamma_{12}t$. Notice that the random vector $MS_2(t)$ consists of two random variables $S_1(t)$ and $S_2(t)$, each of which

has a Skellam distribution. We have,

$$
\begin{aligned}
&\Pr\Big[MS_2(t+h) = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix}\Big] \\
&= \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix}\Big] \\
&= \sum_{i_1=j_1-1}^{j_1+1} \sum_{i_2=j_2-1}^{j_2+1} \left\{ \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ i_2 \end{pmatrix}\Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ i_2 \end{pmatrix}\Big] \right\} + o(h) \\
&= \sum_{i_1=j_1-1}^{j_1+1} \left\{ \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2-1 \end{pmatrix}\Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2-1 \end{pmatrix}\Big] \right. \\
&\quad + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2 \end{pmatrix}\Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2 \end{pmatrix}\Big] \\
&\quad \left. + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2+1 \end{pmatrix}\Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} i_1 \\ j_2+1 \end{pmatrix}\Big] \right\} + o(h).
\end{aligned}
$$

Expanding the second summation, we get

$$
\begin{aligned}
& \Pr\Big[MS_2(t+h) = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big] \\
= \; & \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 - 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 - 1 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 - 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 - 1 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 - 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 - 1 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 + 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 - 1 \\ j_2 + 1 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 + 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 + 1 \end{pmatrix} \Big] \\
& + \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 + 1 \end{pmatrix} \Big] \Pr\Big[\begin{pmatrix} S_1(t+h) \\ S_2(t+h) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big| \begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 + 1 \\ j_2 + 1 \end{pmatrix} \Big] \\
& + o(h)
\end{aligned}
$$

defining the notation

$$
P_{j_1, j_2}(t) = \Pr\Big[\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} j_1 \\ j_2 \end{pmatrix} \Big]
$$

we have,

$$
\begin{aligned}
P_{j_1, j_2}(t+h) \;=\; & \gamma_{12} h P_{j_1 - 1, j_2 - 1}(t) + \lambda_{21} h P_{j_1, j_2 - 1}(t) + \lambda_{11} h P_{j_1 - 1, j_2}(t) + \lambda_{12} h P_{j_1 + 1, j_2}(t) \\
& + \Big[1 - (\lambda_{11} + \lambda_{12} + \lambda_{21} + \lambda_{22} + 2\gamma_{12}) h \Big] P_{j_1, j_2}(t) + \lambda_{22} h P_{j_1, j_2 + 1}(t) \\
& + \gamma_{12} h P_{j_1 + 1, j_2 + 1}(t) + o(h) \,.
\end{aligned}
$$

Taking $P_{j_1,j_2}(t)$ to the other side, deviding by h, and taking the limit as h goes to zero we get

$$\begin{aligned}\frac{d}{dt}P_{j_1,j_2} &= \lambda_{12}\Big(P_{j_1+1,j_2}(t)-P_{j_1,j_2}(t)\Big)\\ &\quad -\lambda_{11}\Big(P_{j_1,j_2}(t)-P_{j_1-1,j_2}(t)\Big)\\ &\quad +\lambda_{22}\Big(P_{j_1,j_2+1}(t)-P_{j_1,j_2}(t)\Big)\\ &\quad -\lambda_{21}\Big(P_{j_1,j_2}(t)-P_{j_1,j_2-1}(t)\Big)\\ &\quad +\gamma_{12}\Big(P_{j_1+1,j_2+1}(t)-P_{j_1,j_2}(t)\Big)\\ &\quad -\gamma_{12}\Big(P_{j_1,j_2}(t)-P_{j_1-1,j_2-1}(t)\Big). \qquad (6.9.16)\end{aligned}$$

This ia bivariate differential-difference equation for $P_{j_1,j_2}(t)$ where the boundary condition is $P_{j_1,j_2}(0)=I_{\{j_1=j_2=0\}}(0)$. Equation (6.9.16) is clearly simplified to the univariate case if $\gamma_{12}=\lambda_{21}=\lambda_{22}=0$. Using the probability generating function method we have

$$\begin{aligned}\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\frac{d}{dt}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2} &= \lambda_{12}\left[\sum_{j_1,j_2}P_{j_1+1,j_2}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}\right]\\ &- \lambda_{11}\left[\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1-1,j_2}P_{j_1-1,j_2}(t)w_1^{j_1}w_2^{j_2}\right]\\ &+ \lambda_{22}\left[\sum_{j_1,j_2}P_{j_1,j_2+1}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}\right]\\ &- \lambda_{21}\left[\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1,j_2}P_{j_1,j_2-1}(t)w_1^{j_1}w_2^{j_2}\right]\\ &+ \gamma_{12}\left[\sum_{j_1,j_2}P_{j_1+1,j_2+1}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}\right]\\ &- \gamma_{12}\left[\sum_{j_1,j_2}P_{j_1,j_2}(t)w_1^{j_1}w_2^{j_2}-\sum_{j_1,j_2}P_{j_1-1,j_2-1}(t)w_1^{j_1}w_2^{j_2}\right],\end{aligned}$$

hence,

$$\begin{aligned}\frac{\partial}{\partial t}G(w_1,w_2,t) &= \lambda_{12}\left[\frac{1}{w_1}G(w1,w2,t)-G(w1,w2,t)\right]-\lambda_{11}\left[G(w_1,w_2,t)-w_1G(w_1,w_2,t)\right]\\ &+ \lambda_{22}\left[\frac{1}{w_2}G(w_1,w_2,t)-G(w_1,w_2,t)\right]-\lambda_{21}\left[G(w_1,w_2,t)-w_2G(w_1,w_2,t)\right]\\ &+ \gamma_{12}\left[\frac{1}{w_1w_2}G(w_1,w_2,t)-G(w_1,w_2,t)\right]\\ &- \gamma_{12}\left[G(w_1,w_2,t)-w_1w_2G(w_1,w_2,t)\right].\end{aligned}$$

The boundary condition is $G(w_1,w_2,0)=1$. So

$$\begin{aligned}\frac{\partial G(w_1,w_2,t)/\partial t}{G(w_1,w_2,t)} &= -\Big(\lambda_{11}+\lambda_{12}+\lambda_{21}+\lambda_{22}+2\gamma_{12}\Big)+\lambda_{11}w_1+\lambda_{21}w_2+\gamma_{12}w_1w_2\\ &\quad +\frac{\lambda_{12}}{w_1}+\frac{\lambda_{22}}{w_2}+\frac{\gamma_{12}}{w_1w_2}\,.\end{aligned}$$

The unique solution to the differential equation above is

$$\begin{aligned}G(w_1,w_2,t) &= \exp\left\{-\Big(\lambda_{11}+\lambda_{12}+\lambda_{21}+\lambda_{22}+2\gamma_{12}\Big)+\lambda_{11}w_1+\lambda_{21}w_2+\gamma_{12}w_1w_2\right.\\ &\quad \left.+\frac{\lambda_{12}}{w_1}+\frac{\lambda_{22}}{w_2}+\frac{\gamma_{12}}{w_1w_2}\right\}.\end{aligned}$$

This is the generating function of a bivariate Skellam distribution with $+,+$ configuration in Equation (6.3.11) as required.

Now we prove the other direction. It suffices to show that for any value $t \geq 0$ and positive infinitesimal h, conditions 3-11 of Definition 6.1 are satisfied. We can show this by writing the bivariate Taylor expansion of the bivariate Skellam distribution about $h=0$ under different values of the vector z listed in conditions 3-11. Another solution, which is much simpler, is through writing the probabilities for the univariate random variables which form the multivariate random vector. Let

$$\underset{h}{\Delta}Y(t) = Y(t+h)-Y(t)\,. \tag{6.9.17}$$

Since

$$Z = \begin{pmatrix} Y_1 + Y_{12} \\ Y_2 + Y_{12} \end{pmatrix}$$

then

$$\underset{h}{\Delta} Z(t) = Z(t+h) - Z(t) = \begin{pmatrix} \underset{h}{\Delta} Y_1(t) + \underset{h}{\Delta} Y_{12}(t) \\ \underset{h}{\Delta} Y_2(t) + \underset{h}{\Delta} Y_{12}(t) \end{pmatrix}. \tag{6.9.18}$$

We show that condition 3 is satisfied;

$$\begin{aligned} \Pr\Big[\underset{h}{\Delta} Z(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix} \underset{h}{\Delta} Y_1(t) + \underset{h}{\Delta} Y_{12}(t) \\ \underset{h}{\Delta} Y_2(t) + \underset{h}{\Delta} Y_{12}(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\Big] \\ &= \Pr\Big(\underset{h}{\Delta} Y_1(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_{12}(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_2(t) = 0\Big) + o(h) \\ &= \Big(1 - \lambda_{11}h - \lambda_{12}h\Big)\Big(1 - \lambda_{21}h - \lambda_{22}h\Big)\Big(1 - 2\gamma_{12}h\Big) + o(h) \\ &= 1 - \Big(\lambda_{11} + \lambda_{12} + \lambda_{21} + \lambda_{22} + 2\gamma_{12}\Big)h + o(h). \end{aligned}$$

Following similar derivations, conditions 4-11 are satisfied.
Condition 4:

$$\begin{aligned} \Pr\Big[\underset{h}{\Delta} Z(t) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix} \underset{h}{\Delta} Y_1(t) + \underset{h}{\Delta} Y_{12}(t) \\ \underset{h}{\Delta} Y_2(t) + \underset{h}{\Delta} Y_{12}(t) \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\Big] \\ &\simeq \Pr\Big(\underset{h}{\Delta} Y_1(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_{12}(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_2(t) = -1\Big) \\ &= \Big(1 - \lambda_{11}h - \lambda_{12}h\Big)\Big(1 - 2\gamma_{12}h\Big)\Big(\lambda_{22}h\Big) + o(h) \\ &= \lambda_{22}h + o(h). \end{aligned}$$

Condition 5:

$$\begin{aligned} \Pr\Big[\underset{h}{\Delta} Z(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix} \underset{h}{\Delta} Y_1(t) + \underset{h}{\Delta} Y_{12}(t) \\ \underset{h}{\Delta} Y_2(t) + \underset{h}{\Delta} Y_{12}(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\Big] \\ &\simeq \Pr\Big(\underset{h}{\Delta} Y_1(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_{12}(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_2(t) = 1\Big) \\ &= \Big(1 - \lambda_{11}h - \lambda_{12}h\Big)\Big(1 - 2\gamma_{12}h\Big)\Big(\lambda_{21}h\Big) + o(h) \\ &= \lambda_{21}h + o(h). \end{aligned}$$

Condition 6:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta}Z(t) = \begin{pmatrix}-1\\0\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta}Y_1(t) + \underset{h}{\Delta}Y_{12}(t)\\ \underset{h}{\Delta}Y_2(t) + \underset{h}{\Delta}Y_{12}(t)\end{pmatrix} = \begin{pmatrix}-1\\0\end{pmatrix}\Big]\\
&\simeq \Pr\Big(\underset{h}{\Delta}Y_1(t) = -1\Big)\Pr\Big(\underset{h}{\Delta}Y_{12}(t) = 0\Big)\Pr\Big(\underset{h}{\Delta}Y_2(t) = 0\Big)\\
&= \Big(\lambda_{12}h\Big)\Big(1 - 2\gamma_{12}h\Big)\Big(1 - \lambda_{21}h - \lambda_{22}h\Big) + o(h)\\
&= \lambda_{12}h + o(h).
\end{aligned}$$

Condition 7:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta}Z(t) = \begin{pmatrix}-1\\-1\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta}Y_1(t) + \underset{h}{\Delta}Y_{12}(t)\\ \underset{h}{\Delta}Y_2(t) + \underset{h}{\Delta}Y_{12}(t)\end{pmatrix} = \begin{pmatrix}-1\\-1\end{pmatrix}\Big]\\
&\simeq \Pr\Big(\underset{h}{\Delta}Y_1(t) = 0\Big)\Pr\Big(\underset{h}{\Delta}Y_{12}(t) = -1\Big)\Pr\Big(\underset{h}{\Delta}Y_2(t) = 0\Big)\\
&= \Big(1 - \lambda_{11}h - \lambda_{12}h\Big)\Big(\gamma_{12}h\Big)\Big(1 - \lambda_{21}h - \lambda_{22}h\Big) + o(h)\\
&= \gamma_{12}h + o(h).
\end{aligned}$$

Condition 8:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta}Z(t) = \begin{pmatrix}-1\\1\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta}Y_1(t) + \underset{h}{\Delta}Y_{12}(t)\\ \underset{h}{\Delta}Y_2(t) + \underset{h}{\Delta}Y_{12}(t)\end{pmatrix} = \begin{pmatrix}-1\\1\end{pmatrix}\Big]\\
&= o(h).
\end{aligned}$$

Condition 9:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta}Z(t) = \begin{pmatrix}1\\0\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta}Y_1(t) + \underset{h}{\Delta}Y_{12}(t)\\ \underset{h}{\Delta}Y_2(t) + \underset{h}{\Delta}Y_{12}(t)\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}\Big]\\
&\simeq \Pr\Big(\underset{h}{\Delta}Y_1(t) = 1\Big)\Pr\Big(\underset{h}{\Delta}Y_{12}(t) = 0\Big)\Pr\Big(\underset{h}{\Delta}Y_2(t) = 0\Big)\\
&= \Big(\lambda_{11}h\Big)\Big(1 - 2\gamma_{12}h\Big)\Big(1 - \lambda_{21}h - \lambda_{22}h\Big) + o(h)\\
&= \lambda_{11}h + o(h).
\end{aligned}$$

Condition 10:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta}Z(t) = \begin{pmatrix}1\\-1\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta}Y_1(t) + \underset{h}{\Delta}Y_{12}(t)\\ \underset{h}{\Delta}Y_2(t) + \underset{h}{\Delta}Y_{12}(t)\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}\Big]\\
&= o(h).
\end{aligned}$$

Condition 11:

$$\begin{aligned}
\Pr\Big[\underset{h}{\Delta} Z(t) = \begin{pmatrix}1\\1\end{pmatrix}\Big] &= \Pr\Big[\begin{pmatrix}\underset{h}{\Delta} Y_1(t) + \underset{h}{\Delta} Y_{12}(t)\\ \underset{h}{\Delta} Y_2(t) + \underset{h}{\Delta} Y_{12}(t)\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}\Big]\\
&\simeq \Pr\Big(\underset{h}{\Delta} Y_1(t) = 0\Big)\Pr\Big(\underset{h}{\Delta} Y_{12}(t) = 1\Big)\Pr\Big(\underset{h}{\Delta} Y_2(t) = 0\Big)\\
&= \Big(1 - \lambda_{11}h\lambda_{12}h\Big)\Big(\gamma_{12}h\Big)\Big(1 - \lambda_{21}h - \lambda_{22}h\Big) + o(h)\\
&= \gamma_{12}h + o(h).
\end{aligned}$$

6.9.2 Proof of Theorem 6.2

Proof: First we show that Definition 6.3 implies Definition 6.4. Since the idea behind the proof is similar to that of Theorem 4.3 in Chapter 4, we provide only a sketch of the proof here. Let $\{MS_{_2}(t),\ t \geq 0\}$ be a inhomogeneous Skellam process defined by Definition 6.3. Likewise the univariate case,

$$MS_{_2}(t+s) - MS_{_2}(s) \sim MSk_{_2}\left(\int_s^{t+s} \lambda_{11}(y)\,dy\ ,\ldots\ , \int_s^{t+s} \gamma_{12}(y)\,dy\right)$$

can be written as

$$MS_{_2}(t+s) - MS_{_2}(s) \sim MSk_{_2}\left(\int_0^{t} \lambda_{11}(y+s)\,dy\ ,\ldots\ , \int_0^{t} \gamma_{12}(y+s)\,dy\right)$$

by assuming that the inhomogeneous bivariate Skellam process starts at time s, i.e. $MS_{_2}(s) = \binom{0}{0}$. Similar to the proof of the Theorem 4.3 in Chapter 4 it can be shown that for the $(+,+)$ configuration,

$$\begin{aligned}
\frac{\partial G(w_1,w_2,t,s)}{\partial t} &= G(w_1,w_2,t,s)\Bigg[-\Big(\lambda_{11}(t+s) + \lambda_{12}(t+s)\\
&\quad +\lambda_{21}(t+s) + \lambda_{22}(t+s) + 2\gamma_{12}(t+s)\Big) + \lambda_{11}(t+s)w_1\\
&\quad +\lambda_{21}(t+s)w_2 + \gamma_{12}(t+s)w_1w_2 + \frac{\lambda_{12}(t+s)}{w_1}\\
&\quad +\frac{\lambda_{22}(t+s)}{w_2} + \frac{\gamma_{12}(t+s)}{w_1w_2}\Bigg],
\end{aligned} \tag{6.9.19}$$

with the boundary condition $G(w_1,w_2,1,1) = 1$, where G is the generating function. The unique solution to the differential equation (6.9.19) is a bivariate function which

is the same as the PGF of the bivariate Skellam random vector $(+,+$ configuration) with intensity functions $\int_s^{t+s} \lambda_{11}(y)\,dy, ..., \int_s^{t+s} \gamma_{12}(y)\,dy$ as required.

To prove the other direction, let

$$m_h^{(i,j)}(t) = \int_t^{t+h} \lambda_{i,j}(y)\,dy \quad = \quad \Lambda_{ij}(t+h) - \Lambda_{ij}(t) \tag{6.9.20}$$

$$\overset{*}{m}_h^{(i,j)}(t) = \int_t^{t+h} \gamma_{i,j}(y)\,dy \quad = \quad \Gamma_{ij}(t+h) - \Gamma_{ij}(t) \tag{6.9.21}$$

Similar to the proof of Theorem 6.1, there are two solutions.

solution #1: The probability mass function of a bivariate Skellam random vector is

$$\begin{aligned}
&\Pr\Big[MS_2(t+h) - MS_2(t) = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\Big] = \\
&\exp\Big\{ -\Big(m_h^{(1,1)}(t) + m_h^{(1,2)}(t) + m_h^{(2,1)}(t) + m_h^{(2,2)}(t) - 2\overset{*}{m}_h^{(1,2)}(t)\Big)\Big\} \\
&\times \sum_{y_{12}=-\infty}^{\infty} \Bigg\{ \Bigg(\sum_{x=0}^{\infty} \frac{[m_h^{(1,1)}(t)]^{z_1-y_{12}+x}[m_h^{(1,2)}(t)]^x}{(z_1-y_{12}+x)!\,x!} \Bigg) \Bigg(\sum_{x=0}^{\infty} \frac{[m_h^{(2,1)}(t)]^{z_2-y_{12}+x}[m_h^{(2,1)}(t)]^x}{(z_2-y_{12}+x)!\,x!} \Bigg) \\
&\qquad \times \Bigg(\sum_{x=0}^{\infty} \frac{[\overset{*}{m}_h^{(2,1)}(t)]^{y_{12}+2x}}{(y_{12}+x)!\,x!} \Bigg) \Bigg\}
\end{aligned} \tag{6.9.22}$$

Now we should write the Taylor expansion of this PMF. Notice that the Taylor expansion of the mean functions introduced in Equations (6.9.20) and (6.9.21) about the point $h = 0$ are

$$\begin{aligned}
m_h^{(i,j)}(t) &= \Lambda_{ij}(t) + \lambda_{ij}(t)h + o(h) - \Lambda_{ij}(t) \\
&= \lambda_{ij}(t)h + o(h) \\
\overset{*}{m}_h^{(i,j)}(t) &= \Gamma_{ij}(t) + \gamma_{ij}(t)h + o(h) - \Gamma_{ij}(t) \\
&= \gamma_{ij}(t)h + o(h)
\end{aligned}$$

Now writing the Taylor expansion of the probability mass function (6.9.22) about the point $h = 0$ for different cases listed in Definition 6.3 complete the proof.

solution #2: We can break the probability terms in conditions 3-11 of Definition 6.3 to probabilities of univariate random variables using similar notations as those of Equations (6.9.17) and 6.9.18, and employ univariate Taylor expansion in each case as we did in the proof of Theorem 6.1.

The proof above was under the condition of positively correlated Skellam random vector $(+,+$ configuration). The proofs for the other three configurations in the bivariate vector are similar.

6.9.3 The recursive parameter estimation algorithm

Similar to the univariate case we will use a recursive function to compute the log-likelihood function. To make the formulation of the problem easy to follow, we will introduce our methodology in the bivariate case,. The multivariate formulation will be an obvious generalization. Discretizing time, we define the observable random variables X^{obs_1} and X^{obs_2} for neurons 1 and 2, respectively.

$$X_i^{obs_j} = \begin{cases} 1 & \text{if a spike form neuron j occurs at time bin } i \\ 0 & \text{otherwise} \end{cases}$$

Similar to Chapter 5, let $\{Z_i^{(j)},\ i = 0, 1, ...\}$ represents the trajectory of the Skellam process with resetting for neuron j. Furthermore, let k be, the number of steps from the current time to the next time where $Z_i^{(1)} = 1$. Analogous to the value k, let ℓ be the same values as k but for the second process. In other words,

$$\begin{aligned} k &:= \min\{i :\ Z_i^{(1)} = 1,\ i > \text{current time}\} \\ \ell &:= \min\{i :\ Z_i^{(2)} = 1,\ i > \text{current time}\} \end{aligned}$$

where $k, \ell \in \{1, 2, 3, ...\}$. The quantity

$$P(k, \ell | m, n) = \Pr\left(X_{k+i}^{obs_1} = 1, X_{\ell+i}^{obs_2} = 1 \middle| Z_i^{(1)} = m, Z_i^{(2)} = n\right) \tag{6.9.23}$$

denotes the probability of observing the next spike form neuron 1 in exactly $k > 0$ and from neuron 2 in exactly $\ell > 0$ steps while the current state for neuron 1's process is $m \leq 1$, and that of neuron 2 is $n \leq 1$. Representing $MS_2(t+h) - MS_2(t)$ by $MS_2(h)$, we introduce the following notation associated with Theorem 6.1.

$$\Pr\left[MS_2(h) = \begin{pmatrix}0\\0\end{pmatrix}\right] = p_{00} \quad , \quad \Pr\left[MS_2(h) = \begin{pmatrix}0\\-1\end{pmatrix}\right] = p_{0-},$$
$$\Pr\left[MS_2(h) = \begin{pmatrix}0\\1\end{pmatrix}\right] = p_{0+} \quad , \quad \Pr\left[MS_2(h) = \begin{pmatrix}-1\\0\end{pmatrix}\right] = p_{-0},$$
$$\Pr\left[MS_2(h) = \begin{pmatrix}-1\\1\end{pmatrix}\right] = p_{-+} \quad , \quad \Pr\left[MS_2(h) = \begin{pmatrix}-1\\-1\end{pmatrix}\right] = p_{--},$$
$$\Pr\left[MS_2(h) = \begin{pmatrix}1\\0\end{pmatrix}\right] = p_{+0} \quad , \quad \Pr\left[MS_2(h) = \begin{pmatrix}1\\-1\end{pmatrix}\right] = p_{+-},$$
$$\Pr\left[MS_2(h) = \begin{pmatrix}1\\1\end{pmatrix}\right] = p_{++}. \tag{6.9.24}$$

Now, conditioning $P(k,\ell|m,n)$ in Equation 6.9.23 on the first step of each process we have

If $m < 0$ and $n < 0$,

$$\begin{aligned} P(k,\ell|m,n) &= p_{--}P(k-1,\ell-1|m-1,n-1) + p_{0-}P(k-1,\ell-1|m,n-1) \\ &\quad + p_{+-}P(k-1,\ell-1|m+1,n-1) + p_{-0}P(k-1,\ell-1|m-1,n) \\ &\quad + p_{00}P(k-1,\ell-1|m,n) + p_{+0}P(k-1,\ell-1|m+1,n) \\ &\quad + p_{-+}P(k-1,\ell-1|m-1,n+1) + p_{0+}P(k-1,\ell-1|m,n+1) \\ &\quad + p_{++}P(k-1,\ell-1|m+1,n+1) \end{aligned} \tag{6.9.25}$$

If $m < 0$ and $n = 0$,

$$\begin{aligned} P(k,\ell|m,0) &= p_{--}P(k-1,\ell-1|m-1,-1) + p_{0-}P(k-1,\ell-1|m,-1) \\ &\quad + p_{+-}P(k-1,\ell-1|m+1,-1) + p_{-0}P(k-1,\ell-1|m-1,0) \\ &\quad + p_{00}P(k-1,\ell-1|m,0) + p_{+0}P(k-1,\ell-1|m+1,0) \end{aligned} \tag{6.9.26}$$

If $m = 0$ and $n < 0$,

$$\begin{aligned} P(k,\ell|0,n) &= p_{--}P(k-1,\ell-1|-1,n-1) + p_{0-}P(k-1,\ell-1|0,n-1) \\ &\quad + p_{-0}P(k-1,\ell-1|-1,n) + p_{00}P(k-1,\ell-1|0,n) \\ &\quad + p_{-+}P(k-1,\ell-1|-1,n+1) + p_{0+}P(k-1,\ell-1|0,n+1) \end{aligned} \tag{6.9.27}$$

If $m = 0$ and $n = 0$,

$$\begin{aligned} P(k,\ell|0,0) &= p_{--}P(k-1,\ell-1|-1,-1) + p_{0-}P(k-1,\ell-1|0,-1) \\ & \quad p_{-0}P(k-1,\ell-1|-1,0) + p_{00}P(k-1,\ell-1|0,0) \end{aligned} \tag{6.9.28}$$

The initial/boundary conditions for Equations (6.9.25)-(6.9.28) are

$$P(k,\ell|m,n) = \begin{cases} 0 & \text{if } (m \le -k) \text{ or } (n \le -\ell) \text{ or } (m = k = 1) \text{ or } (n = \ell = 1) \\ p_{++} & \text{if } (k = 1 \ \& \ m = 0) \text{ and } (\ell = 1 \ \& \ n = 0) \\ A & \text{if } (k = 1 \ \& \ m = 0) \text{ and } n = 1 \\ B & \text{if } m = 1 \text{ and } (\ell = 1 \ \& \ n = 0) \\ P(k-1,\ell-1|0,0) & \text{if } m = 1 \text{ and } n = 1 \end{cases}$$

where

$$\begin{aligned} A &= P(1,\ell-1|0,0) \\ &= p_{++}I(\ell-1=1) \\ &\quad + \Big[p_{+0}P(k,\ell-2|m,0) + p_{+-}P(k,\ell-2|m,-1)\Big] I(\ell-1 \ge 1)\,, \end{aligned}$$

and

$$\begin{aligned} B &= P(k-1,1|0,0) \\ &= p_{++}I(k-1=1) \\ &\quad + \Big[p_{0+}P(k-2,\ell|0,n) + p_{-+}P(k-2,\ell|-1,n)\Big] I(k-1 \ge 1)\,. \end{aligned}$$

In the univariate case of Chapter 5, we employed the conditional independence assumption every time a spike was initiated; however, that assumption should be adjusted for the bivariate case as resettings don't occur simultaneously for different neurons. In this case, whenever one of the neurons fires, we use the conditional independence property, but also average over the possible states of the other neuron. Let us show the likelihood derivation in an example. Let $\{t_i^{(j)}, i = 1, 2, ..., n_1\}$ be the spike times of the j^{th} neuron, $j = 1, 2$. Figure 6.4 shows a possible permutation of the spike times and the number of steps between consecutive spikes for each neuron. The two quantities k_i and ℓ_j in this graph show the number of steps between consecutive spikes.

We will derive the likelihood function for this example. Likelihood derivation based on any other permutation of spikes is similar. Let us derive the likelihood terms corresponding to the intervals before the first and after the last spikes. Similar to the

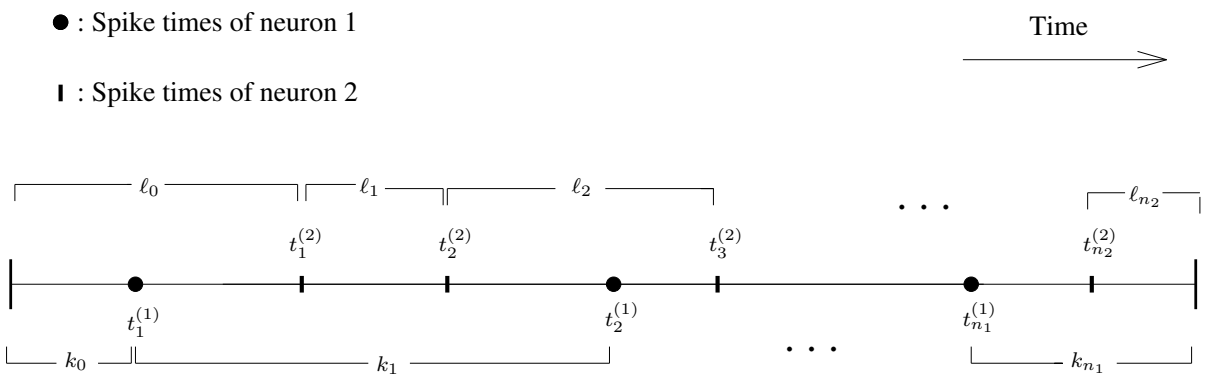


Figure 6.4: *One example of possible permutations of the spikes from two neurons N_1 and N_2.*

univariate case, we make use of stationary distributions of the marginal processes. Let $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ be the stationary distributions of the SPRs associated with neuron 1 and neuron 2, respectively. First, consider the interval before the very first spike, i.e. $\big[0\,,t_1^{(1)}\big)$. Define the events $A_{k_0} := \{X_j^{obs_1} = X_j^{obs_2} = 0 \text{ for } j < k_0\}$ and $B_{\ell_0} := \{X_{j_2}^{obs_2} = 0 \text{ for } j_2 < \ell_0\}$. We have,

$$
\begin{aligned}
& \Pr\Big(X_{k_0}^{obs_1} = 1\ ,\ X_{\ell_0}^{obs_2} = 1\ ,\ A_{k_0}\ ,\ B_{\ell_0}\Big) \\
= & \sum_m \sum_n \Pr\Big(X_{k_0}^{obs_1} = X_{\ell_0}^{obs_2} = 1\,, A_{k_0}, B_{\ell_0}\Big| Z_0^{(1)} = m\,, Z_0^{(2)} = n\Big)\Pr\Big(Z_0^{(1)} = m\,, Z_0^{(2)} = n\Big) \\
= & \sum_m \sum_n P(k_0, \ell_0 | m, n) \Pr\Big(Z_0^{(1)} = m, Z_0^{(2)} = n\Big)
\end{aligned}
$$

We substitute the term $\Pr\big(Z_0^{(1)} = m, Z_0^{(2)} = n\big)$ with the product of marginal stationary distributions $\pi_{2-m}^{(1)}\,\pi_{2-n}^{(2)}$. Therefore,

$$
\Pr\Big(X_{k_0}^{obs_1} = 1\ ,\ X_{\ell_0}^{obs_2} = 1\ ,\ A_{k_0}\ ,\ B_{\ell_0}\Big) = \sum_m \sum_n P(k_0, \ell_0 | m, n)\, \pi_{2-m}^{(1)}\, \pi_{2-n}^{(2)} \quad (6.9.29)
$$

Now, we derive the terms for the periods after the last spikes. According to Figure 6.4, neuron 2 fires the las spike at time $t_{n_2}^{(2)}$, therefore, averaging over all possible values of the trajectory of the other neurons' SPR, for the time bin after the very last spike in Figure 6.4 we have

$$\Pr\Big(X_{1+k^*}^{obs_i}=0,\ X_{2+k^*}^{obs_i}=0,\quad \ldots\quad ,X_{k_{n_1}+k^*}^{obs_i}=0\ \text{ for } i=1,2\ \Big|\ Z_{k^*}=1\Big)=$$
$$\sum_m\left\{\left[1-\sum_{j_1=1}^{\ell_{n_2}}\sum_{j_2=1}^{\ell_{n_2}}P\Big(j_1\,,j_2\Big|m\,,1\Big)\right]\Pr\Big(Z_{t_{n_2}^{(2)}}^{(1)}=m\Big)\right\} \tag{6.9.30}$$

where $k^*=\sum_{i=0}^{n_2-1}\ell_i$. Similarly, corresponding to the period between the last spike of the first neuron and and the last spike of the second neuron, i.e. $\big[t_{n_1}^{(1)}\,,t_{n_2}^{(2)}\big)$, we have

$$\Pr\Big(X_{1+k^{**}}^{obs_1}=0,\ X_{2+k^{**}}^{obs_1}=0,\ \ldots,X_{w+k^{**}}^{obs_1}=0\ \text{ for } i=1,2\ \Big|\ Z_{k^*}=1\Big)=$$
$$\sum_n\left\{\left[1-\sum_{j=1}^{k_n}P\Big(j\,,\sum_{i=0}^{n_2-1}\ell_i-\sum_{i=0}^{n_1-1}k_i\Big|1\,,n\Big)\right]\Pr\Big(Z_{t_{n_1}^{(1)}}^{(2)}=n\Big)\right\} \tag{6.9.31}$$

where $k^{**}=\sum_{i=0}^{n_1-1}k_i$ and $w=\sum_{i=0}^{n_2-1}\ell_i+\sum_{i=0}^{n_1-1}k_i$. We substitute $\Pr\big(Z_{t_{n_1}^{(1)}}^{(2)}=n\big)$ and $\Pr\big(Z_{t_{n_1}^{(2)}}^{(1)}=m\big)$ with the marginal stationary distributions $\pi_{2-n}^{(2)}$ and $\pi_{2-m}^{(1)}$, respectively. Equations (6.9.29), (6.9.30) and (6.9.31) count for the contribution of the first and last subintervals of the spiking activity. The full likelihood function is

$$\begin{aligned}
L\Big(\mathbf{p};\boldsymbol{x}^{obs_1},\boldsymbol{x}^{obs_2}\Big) \;=\; & \sum_{m=-k_0+1}^{1}\sum_{n=-\ell_0+1}^{1}P\Big(k_0\,,\ell_0\Big|m\,,n\Big)\pi_{2-m}^{(1)}\,\pi_{2-n}^{(2)}\times\\
& \sum_n P\Big(k_1\,,\ell_0-k_0\Big|1\,,n\Big)\,\pi_{2-n}^{(2)}\times\\
& \sum_m P\Big(\sum_{i=0}^{1}k_i-\ell_0\,,\ell_1\Big|m\,,1\Big)\,\pi_{2-m}^{(1)}\times\\
& \sum_m P\Big(\sum_{i=0}^{1}k_i-\sum_{i=0}^{1}\ell_i\,,\ell_2\Big|m\,,1\Big)\,\pi_{2-m}^{(1)}\times\\
& \sum_n P\Big(k_2\,,\sum_{i=0}^{2}\ell_i-\sum_{i=0}^{1}k_i\Big|1\,,n\Big)\,\pi_{2-n}^{(2)}\times\cdots\times\\
& \sum_n\left\{\left[1-\sum_{j=1}^{k_{n_1}-\ell_{n_2}}P\Big(j\,,\sum_{i=0}^{n_2-1}\ell_i-\sum_{i=0}^{n_1-1}k_i\Big|1\,,n\Big)\right]\pi_{2-n}^{(2)}\right\}\times\\
& \sum_m\left\{\left[1-\sum_{j_1=1}^{\ell_{n_2}}\sum_{j_2=1}^{\ell_{n_2}}P\Big(j_1\,,j_2\Big|m\,,1\Big)\right]\pi_{2-m}^{(1)}\right\}
\end{aligned} \tag{6.9.32}$$

where $\mathbf{p} = (p_{--}, p_{-+}, ..., p_{++})$ is the vector of parameters. Taking the logarithm from both sides of Equation (6.9.32) we get the loglikelihood function of the bivariate spike trains,

$$\ell\big(\mathbf{p}; \boldsymbol{x}^{obs_1}, \boldsymbol{x}^{obs_2}\big) = \log\Big(L\big(\mathbf{p}; \boldsymbol{x}^{obs_1}, \boldsymbol{x}^{obs_2}\big)\Big).$$

Chapter 7

Discussion and Future Work

7.1 Introduction

This thesis focused on the statistical analysis of neural spike trains. We started with an introduction to the anatomy and physiology of the nervous system and the concept of information. Combining the work of Bickel et al. (2007, 2008) and Shao and Lii (2011) with the multiscale approach of Kolaczyk and Nowak (2004), additive and multiplicative multiscale models for the intensity function of neural spike trains within the framework of inhomogeneous Poisson process were introduced and studied in detail. To address some of the issues related to Poisson process framework, Skellam process with resetting (SPR) was introduced and its theoretical properties were discussed. This model was then generalized to the multivariate case to address the challenging problem of simultaneous analysis of spike trains from multiple neurons. In both univariate and multivariate cases, computationally efficient parameter estimation methods were developed. In this chapter we discuss the "next steps" of the work laid out in earlier chapters.

7.2 Multiscale analysis within Poisson framework

The details of the multiscale analysis of neural spike trains were discussed in chapter 2 and 3. The two additive and multiplicative intensity functions with periodic

components proposed there are

$$\theta_m(t) = c(t)\Big\{(1 - \sum_{k=1}^{K} \eta_k) + \sum_{k=1}^{K} \eta_k \, \nu_{\gamma_k}\big(f_k t + \omega_k^{(0)}\big)\Big\}, \tag{7.2.1}$$

$$\theta_a(t) = \left(1 - \sum_{k=1}^{K} \eta_k\right) c(t) + \sum_{k=1}^{K} \eta_k \nu_{\gamma_k}\big(f_k t + \omega_k^{(0)}\big), \tag{7.2.2}$$

where

$$\nu_{\gamma_k}(x) = \gamma_k \left[1 + \cos(2\pi\, x)\right], \qquad \text{for} \quad \omega_k^{(0)} < x < f_k\, T + \omega_k^{(0)}.$$

The models discussed in Chapter 2 assume that the initial phases $\boldsymbol{\omega}^{(\mathbf{0})}$ are constant throughout the study period $[0, T)$. This model can be useful for the firing activity of phase-locked neurons, Taniguchi and Ogawa (1987); Köppl (1997); Simoes et al. (2003). Notice that the notion of phase can be used differently across neuroscience literature. We emphasize that here the term "phase" simply refers to $\boldsymbol{\omega}^{(0)}$. The initial phase can also change over time. Some evidence which suggests phase synchronization as one of the important methods of functional integration in the brain is reviewed in Varela et al. (2001). Furthermore, O'Keefe and Recce (1993) discovered the *phase precession* phenomenon which is a property of hippocampal place cells of moving rats. They show that the phase of the spikes from the place cells relative to the θ-rhythm of the local field potential (LFP) decreases as the rat progresses through the field in which it moves. Although their definition of phase is slightly different than that of ours, dynamic changes in one, result in changes in values of the other. These pieces of evidence suggest temporal estimation of $\boldsymbol{\omega}^{(0)}$. One temporal estimate of $\boldsymbol{\omega}^{(0)}$ could be the multiscale estimate, which allows for different values for disjoint time intervals. If the instantaneous phase of the whole intensity function is of interest, then the Hilbert transform is a powerful tool to employ. However, if interest lies in individual phases, which is usually the case in the analysis of brain rhythms, then we suggest keeping several initial phases in the model. Studying this problem is left for future work.

7.3 Univariate Skellam process with resetting (SPR)

It was shown in Chapter 4 that the Skellam model is valid for over-dispersed data (relative to Poisson). The fundamental question is if it is possible to develop a variation of Skellam process such that it can accommodate under-dispersion. We have noticed, through simulations, that such variation does exist. Discretizing time into bins of size h, where $h > 0$ is an infinitesimal, one can simulate a Skellam process with resetting parametrized by λ_1 and λ_2 in a random walk fashion as follows;

1. Define $p_{+} = \lambda_1 h$, $p_{-} = \lambda_2 h$, $p_0 = 1 - (\lambda_1 + \lambda_2)h$, and set $S(0) = 0$ and $t = 1$.

2. Calculate $S(t) = S(t-1) + w$, where w is a number in $\{-1, 0, +1\}$ chosen randomly with respect to probabilities $P(-1) = p_{-}$, $P(0) = p_0$ and $P(+1) = p_{+}$.

3. If $\sum_{i=0}^{t} S(i) = 1$ then set $S(t+1) = 0$, $t \leftarrow t+2$ and go to step 2. Otherwise, $t \leftarrow t+1$ and go to step 2.

The sequence S forms a sample path of the SPR with parameters λ_1 and λ_2. In our simulations, the realizations $S(t) = 1$ are associated with spike time, and counting the number of spikes in time bins of length T results in spike count data. Notice that the important point in this simulation is that $h > 0$ is a small number. In fact, relative to λ_1 and λ_2, h should be negligible. However, if h is not relatively small, we have noticed that this process generates under dispersed data. Figure 7.1 plots the mean and variance of spike counts against each other. Each point in this graph plots the mean and variance of spike counts in 50 realizations of an SPR with parameters $\lambda_1 \in \{1, 2, ..., 40\}$ and $\lambda_2 \in \{1, 2, ..., 40\}$. The Process has been observed over $[0, T] = [0, 5]$ and the bin size for spike counting is 0.5, also $h = 0.01$. Notice that as λ_1 and/or λ_2 increase, the probabilities p_{-}, p_0 and p_{+} increase drastically.

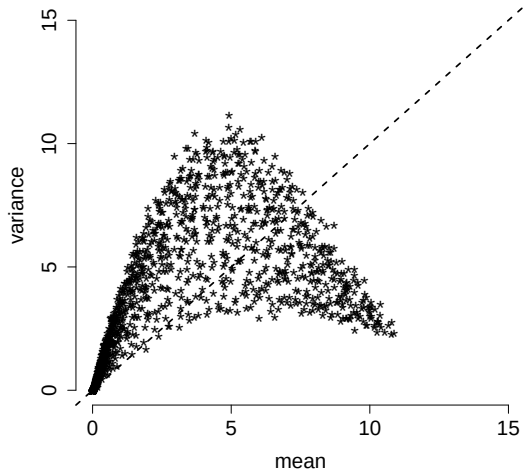


Figure 7.1: *The mean-variance relationship in an SPR model with parameters* $\lambda_1 \in \{1, 2, ..., 40\}$ *and* $\lambda_2 \in \{1, 2, ..., 40\}$ *where* $h = 0.01$. *The dashed line shows the mean=variance line.*

It is clear from the graph that increasing λ_i, $i = 1, 2$ relative to h, causes this under-dispersion effect. As a future research problem, it will be interesting to investigate this model in more details and use it as an approach for modelling under-dispersed spike counts.

Based on Equation (4.5.15) in Chapter 4, it is easy to derive the exact distribution of spike counts. Let Y_i show the i^{th} discretized interspike interval. Furthermore, let $N^*(t)$ denotes the number of spikes in the interval $[0, t]$. It can be shown (see Section 7.5) that the probability mass function of $N^*(t)$ is

$$\Pr(N^*(t) = n) = \sum_{x=\lceil t/h \rceil}^{\infty} \frac{\partial^x \Phi_{n+1}(t)}{\partial t^x}\Big|_{t=0} \times \frac{1}{x!} - \sum_{x=\lceil t/h \rceil}^{\infty} \frac{\partial^x \Phi_n(t)}{\partial t^x}\Big|_{t=0} \times \frac{1}{x!}, \qquad (7.3.3)$$

where

$$\Phi_k(t) = \left[\frac{(1 - tp_0) - \sqrt{(1 - tp_0)^2 - 4t^2 p_- p_+}}{2p_-}\right]^k.$$

We have noticed through simulations that this distribution under time scales that are neither as small or as large as those of Section 4.5 can be well-approximated by the conditional Skellam distribution $\Pr(X|X > 0)$, where $X \sim Sk(\lambda_1, \lambda_2)$. For this simulation study, we generated 1000 realizations of the process with parameters $(\lambda_1, \lambda_2) = (10, 9)$. In this simulation, the length of each step is $h = 0.01$ and $[0, T] = [0, 5]$. Notice that these values of λ_i, $i = 1, 2$ do not satisfy the conditions discussed in Section 4.5. Figure 7.2 shows the raster plot of the simulated data. Notice that the Skellam process was reset to the previous record value after each spike.

The results of this simulation study are summarized in Figure 7.3. The red curve superimposed on the histogram of spike counts, panel (b), is the density function of the Skellam random variable X conditional on $X > 0$. The plot of the data quantiles versus the quantiles of the conditional Skellam distribution, panel (c), confirms the distribution of the data is well approximated by the conditional Skellam distribution. Panel (d) plots the value of the path of the Skellam process with resetting (resetting to the previous record) at $t = 5$ (end of the observation window) versus the spike counts for each of the 1000 trials. The plot shows that the two random variables are positively correlated, but, obviously, are not the same.

7.4 Multivariate Skellam process with resetting (MSPR)

The multivariate Skellam process with resetting (MSPR) was discussed in Chapter 6. As a topic which has just been introduced, there are quite a few places for future work,

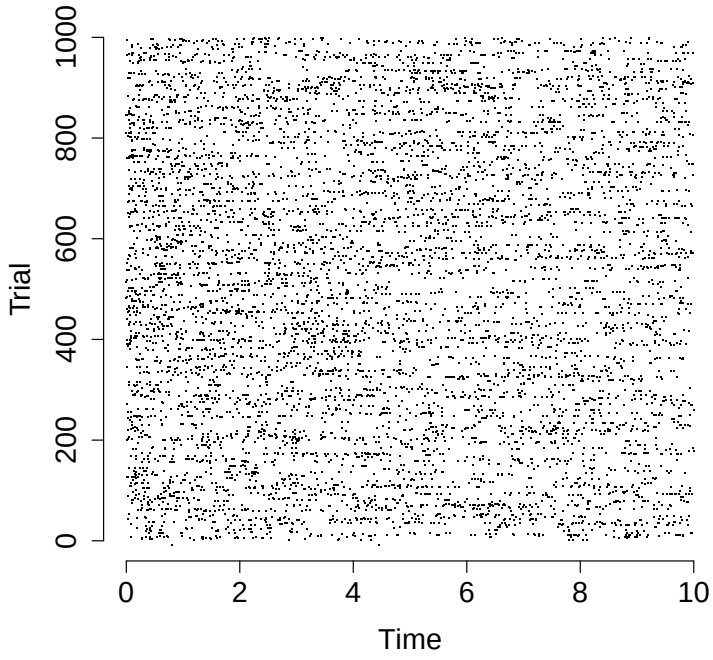


Figure 7.2: *Raster plot of 1000 spike trains generated from the Skellam process with resetting with parameters* $(\lambda_1, \lambda_2) = (10, 9)$.

from both theoretical and computational points of view.

The so-called moment-profile likelihood parameter estimation method was discussed in Chapter 6. It was mentioned, during the real data analysis in that chapter, that using the plug-in moment estimate of parameter γ_{ij} (or the covariances) in the likelihood function may cause bias in the estimation. The effect of using the moment estimate as a plug-in estimate in this particular problem has not been investigated. It is interesting to study the asymptotic behavior of the moment-profile likelihood estimator, and see if it is possible to improve it in terms of both variability and bias (if it is not unbiased).

One problem of interest in the simultaneous analysis of multiple neural spike trains is dimensionality reduction. Analyzing an ensemble of neurons, estimation of the covariance matrix Σ will be a high dimensional problem. Dimensionality reduction, or clustering the neurons is an interesting problem for future work in this area. Based on a collection of p neurons, the first step would be to perform a principal components

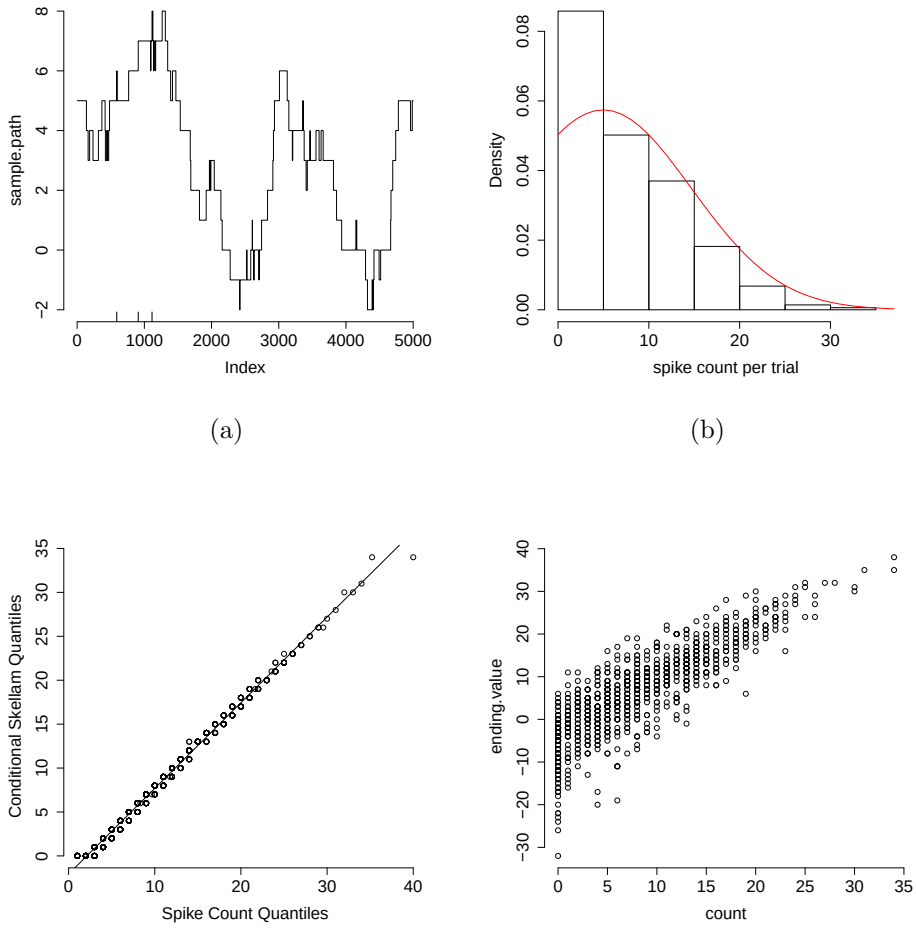


Figure 7.3: *Simulation study based on 1000 realizations of a Skellam process with resetting (to the previous record value) over the interval* $[0, T] = [0, 5]$ *with* $(\lambda_1, \lambda_2) = (10, 9)$. *a: Sample path of one realization. b: The histogram of spike counts per trial with the density of the conditional Skellam distribution superimposed (the red curve). c: The quantiles of the conditional Skellam distribution vs. the quantiles of the data. d: The plot of spike counts per trial vs the value of the Skellam path at the end of interval, i.e. at* $t = 5$.

analysis (PCA) on the $p \times p$ covariance matrix of the spike counts. However, being interested in marginal Skellam models for individual neurons, the optimization problem should be of the form

$$\max_{a} \left(a^T \Sigma a \right) \quad s.t.\ a_i \in \mathbb{Z}\,,$$

where Σ is the covariance matrix of the neural spike trains and the vector $a = (a_1, ..., a_p)^T$ includes the coefficients. The set of integers are shown by $\mathbb{Z}$. Clearly, a_is have to be bounded. We are currently investigating this problem.

Another important, and interesting problem for future work on MSPR is the estimation of the functions $\lambda_{11}(t)$, $\lambda_{12}(t)$, $\lambda_{21}(t)$, $\lambda_{22}(t)$ and $\gamma_{12}(t)$ in the inhomogeneous case. The importance of the multiscale estimation in this context has been discussed in details in Chapters 2 and 3, so generalizing this multiscale approach to the multivariate case will be a valuable estimation method.

We mentioned that the recursive algorithm of Chapter 5 for parameter estimation is not pragmatic in the multivariate case as it is computationally intensive. Further investigation of this algorithm to improve it in terms of speed is an interesting research problem.

7.5 Derivation of spike count distribution

In this appendix we derive the distribution of the spike counts introduced in Equation (7.3.3). We showed in Chapter 4 that

$$\begin{aligned} \Phi(t) &= E(t^{Y_i}) \\ &= \Phi(t) = \frac{(1-tp_0) - \sqrt{(1-tp_0)^2 - 4t^2 p_- p_+}}{2p_-}. \end{aligned}$$

Let T_i show the time of the i^{th} spike (continuous time scale). Since $T_n = T_1 + (T_2 - T_1) + (T_3 - T_2) + \cdots + (T_n - T_{n-1})$, the PGF of the discretized interspike interval is

$$\Phi_n(t) = \left[\frac{(1-tp_0) - \sqrt{(1-tp_0)^2 - 4t^2 p_- p_+}}{2p_-}\right]^n.$$

Now, we have

$$\begin{aligned} \Pr\Big(N^*(t) < n\Big) &= \Pr\Big(T_n > t\Big) \\ &= \Pr\Big((\sum_{i=1}^{n} Y_i)h > t\Big) \\ &= \Pr\Big(\sum_{i=1}^{n} Y_i > t/h\Big) \end{aligned}$$

However,

$$\Pr\Big(\sum_{i=1}^{n} Y_i = k\Big) = \frac{\partial^k \Phi_n(t)}{\partial t^k}\Big|_{t=0} \times \frac{1}{k!}\,,$$

therefore,

$$\begin{aligned}
\Pr\Big(N^*(t) = n\Big) &= \Pr\Big(N^*(t) < n+1\Big) - \Pr\Big(N^*(t) < n\Big) \\
&= \Pr\Big(\sum_{i=1}^{n+1} Y_i > t/h\Big) - \Pr\Big(\sum_{i=1}^{n} Y_i > t/h\Big) \\
&= \sum_{x=\lceil t/h \rceil}^{\infty} \frac{\partial^x \Phi_{n+1}(t)}{\partial t^x}\Big|_{t=0} \times \frac{1}{x!} - \sum_{x=\lceil t/h \rceil}^{\infty} \frac{\partial^x \Phi_n(t)}{\partial t^x}\Big|_{t=0} \times \frac{1}{x!}\,,
\end{aligned}$$

where

$$\Phi_k(t) = \left[\frac{(1-tp_{_0}) - \sqrt{(1-tp_{_0})^2 - 4t^2 p_{_-} p_{_+}}}{2p_{_-}}\right]^k.$$

Bibliography

Abramowitz, M. and Stegun, I. A., editors (1972). *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*. United States Government Printing Office, Washington, DC.

Adrian, E. D. (1926a). The impulses produced by sensory nerve endings part 1. *Journal of Physiology*, 61(1):49–72.

Adrian, E. D. (1926b). The impulses produced by sensory nerve endings part 4: Impulses from pain receptors. *Journal of Physiology*, 62(1):33–51.

Adrian, E. D. (1928). *The basis of sensation: the action of the sense organs*. W.W. Norton & Co., New York, NY.

Adrian, E. D. (1930). The effects of injury on mammalian nerve fibers. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 106(747):596–618.

Adrian, E. D. (1965). *Nobel Lectures, Physiology or Medicine: 1922-1941*. World Scientific, River Edge, NJ.

Adrian, E. D. and Zotterman, Y. (1926a). The impulses produced by sensory nerve endings part 2: The response of a single end organ. *Journal of Physiology*, 61(2):151–171.

Adrian, E. D. and Zotterman, Y. (1926b). The impulses produced by sensory nerve endings part 3: Impulses set up by touch and pressure. *Journal of Physiology*, 61(4):465–483.

Ahmadian, Y., Pillow, J., Shlens, J., Simoncelli, E., Chichilnisky, E., and Paninski, L. (2009). A decoder-based spike train metric for analyzing the neural code in the retina. In *Computational and systems neuroscience*, Frontiers in Systems Neuroscience.

Akaike, H. (1971). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, pages 267–281. Akademia Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Alzaid, A. A. and Omair, M. A. (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 33(1):17–45.

Ananthanarayanan, R. and Modha, D. S. (2007). Anatomy of cortical simulator. In *SC07 proceeding of the 2007 ACM/IEEE conference on Supercomputing*.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. John Wiley & Sons, New York, NY.

Antic, S., Cohen, L. B., Lam, Y. W., Wachowiak, M., Zecevic, D., and Zochowski, M. (1999). Fast multisite optical measurement of membrane potential: three examples. *The Journal of the Federation of American Societies for Experimental Biology*, 13:s271–s276.

Aoyama, K., Shimizu, K., and Ong, S. H. (2008). A first-passage time random walk distribution with five transition probabilities: a generalization of the shifted inverse trinomial. *Annals of the Institute of Statistical Mathematics*, 60(1):1–20.

Arrillaqa, J. and Watson, N. R. (2003). *Power System Harmonics*. John Wiley & Sons, New York, NY, 2nd edition.

Arzoumanian, Z., Cordes, J. M., and Wasserman, I. (1999). Pulsar spin evolution, kinematics, and the birthrate of neutron star binaries. *The Astrophysical Journal*, 520:696–705.

Averbeck, B. B. (2009). Poisson or not Poisson: differences in spike train statistics between parietal cortical areas. *Neuron*, 62(3):310–311.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7:358–366.

Axmacher, N., Henseler, M., Jensen, O., Weinreich, I., Elger, C., and Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *PNAS*, 10(7):3228–3233.

Bachman, G., Lawrence, N., and Beckenstein, E. (2000). *Fourier and Wavelet Analysis*. Springer, New York, NY.

Bai, O., Lin, P., Vorbach, S., Floeter, M. K., Hattori, N., and Hallett, M. (2008). A high performance sensorimotor beta rhythm-based brain-computer interface associated with human natural motor behavior. *Journal of Neural Engineering*, 5:24–35.

Bajaj, C. (1999). *Data Visualization Techniques*. John Wiley & Sons, New York, NY.

Banerjee, A., Dean, H. L., and Pesaran, B. (2012). Parametric models to relate spike train and LFP dynamics with neural information processing. *Frontiers in Computational Neuroscience*, 6:51.

Bar-Gad, I., Ritov, Y., and Bergman, H. (2001). The neuronal refractory period causes a short-term peak in the autocorrelation function. *Journal of Neuroscience Methods*, 104:155–163.

Bar-Lev, S. K. and Pommeret, D. (2003). A note on natural exponential families with cuts. *Statistics and Probability Letters*, 63:215–221.

Barbieri, R., Frank, L., Nguyen, D. P., Quirk, M. C., Solo, V., Wilke, M., and Brown, E. N. (2004). Dynamic analysis of information encoding in neural ensembles. *Neural Computation*, 16:277–307.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., and Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience*, 105:25–37.

Barndorff-Nielsen, O. E. (1976). Factorization of likelihood functions for full exponential families. *Journal of the Royal Statistical Society. Series B*, 38(1).

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, New York, NY.

Barndorff-Nielsen, O. E., Pollard, D. G., and Shephard, N. (2010). Discrete-valued Lévy processes and low latency financial econometrics. Technical Report 490, University of Oxford.

Barndorff-Nielsen, O. E., Pollard, D. G., and Shephard, N. (2012). Integer-valued Lévy processes and low latency financial econometrics. *Quantitative Finance*, 4:587–605.

Barreto, A., Chin, N., Andrian, J., and Riley, J. (1995). Multi-resolution characterization of interictal epileptic spikes based on wavelet transformation. In *Biomedical Engineering Conference*.

Bartlett, M. S. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):264–296.

Bartlett, M. S. (1964). The spectral analysis of two dimensional point processes. *Biometrika*, 51(3/4):299–311.

Bartlett, M. S. (1978). *An Introduction to Stochastic Processes*. Cambridge University Press, Cambridge, United Kingdom, third edition.

Beck, H., Goussakov, I. V., Lie, A., Helmstaedter, C., and Elger, C. E. (2000). Synaptic plasticity in the human dentate gyrus. *The Journal of Neuroscience*, 20(18):7080–7086.

Behseta, S. and Chenouri, S. (2011). Comparison of two population of curves with an application in neuronal data analysis. *Statistics in Medicine*, 30(12):1441–54.

Belluscio, M., Mizuseki, K., Schmidt, R., Kempter, R., and Buzsáki, G. (2012). Cross-frequency phase-phase coupling between theta and gamma oscillations in the hippocampus. *The Journal of Neuroscience*, 32(2):423–435.

Beran, R. J. (1969). Asymptotic theory of a class of tests for uniformity of a circular distribution. *The Annals of Mathematical Statistics*, 40(4):1196–1206.

Beutler, F. J. and Leneman, O. A. Z. (1968). The spectral analysis of impulse processes. *Information and Control*, 12:136–158.

Bickel, P., Kleijin, B., and Rice, J. (2007). On detecting periodicity in astronomical point processes. *Statistical Challenges in Modern Astronomy IV. ASP Conference Series*, 371:305–319.

Bickel, P., Kleijin, B., and Rice, J. (2008). Event weighted tests for detecting periodicity in photon arrival times. *The Astrophysical Journal*, 685:384–389.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.

Bird, C. M. and Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, 9:182–194.

Blum, A. S. and Rutkove, S. B., editors (2007). *The Clinical Neurophysiology Primer*. Humana Press, Totowa, NJ.

Blumer, D. and Benson, D. (1975). Personality changes with frontal and temporal lobe lesions. In *Psychiatric Aspects of Neurologic Disease*. Grune and Stratton.

Bonmassar, G., Purdon, P. L., Jääskeläinen, I. P., Chippa, K., Solo, V., Brown, E. N., and Belliveau, J. W. (2002). Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI. *Neuroimage*, 16:1127–1141.

Borg-Graham, L. J. (1999). Interpretations of data and mechanisms for hippocampal pyramidal cell models. In Ulinski, P. S., Jones, E. G., and Peters, A., editors, *Cortical Models*, volume 13 of *Cerebral Cortex*, chapter 13, pages 19–138. New York: Plenum.

Borisyuk, R. M. and Borisyuk, G. N. (1997). Information coding on the basis of synchronization of neuronal activity. *Biosystems*, 40(1-2):2–10.

Bressler, S. L., Coppola, R., and Nakamura, R. (1993). Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature*, 366(11):153–156.

Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59:189–200.

Brillinger, D. R. (2001). *Time series: data analysis and theory*. SIAM: Society for Industrial and Applied Mathematics, London, United Kingdom.

Brink, F., Bronk, D. W., and Larrabee, M. G. (1946). Chemical excitation of nerve. *Annals of the New York Academy of Sciences*, 47:457–487.

Brockwell, A. E., Rojas, A. L., and Kass, R. E. (2004). Recursive bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, 91:1899–1907.

Brown, E., Barbieri, R., Ventura, V., Kass, R., and Frank, L. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346.

Brown, E., Kass, R. E., and Miller, J. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience*, 7(5):456–461.

Brown, E. N., Barbieri, R., Eden, U. T., and Frank, L. M. (2003). Likelihood methods for neural spike train data analysis. In Feng, J., editor, *Computational Neuroscience: A comprehensive approach*. CRC Press.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18):7411–7425.

Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., and Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Neurobiology*, 98(21):12261–12266.

Brown, K. T. and Flaming, D. G. (1986). *Advanced micropipette techniques for cell physiology*. John Wiley & Sons, London, New York, NY.

Brunel, N. (2010). Modeling point neurons: From hodgkin-huxley to integrate-and-fire. In De Schutter, E., editor, *Computational Modeling Methods for Neuroscientists*. MIT Press.

Brunel, N., Chance, F., Fourcaud, N., and Abbott, L. F. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters*, 86(10):2186–2189.

Buccheri, R., Bennett, K., Bignami, G. F., Bloemen, J. B. G. M., Boriakoff, V., Caraveo, P. A., Hermsen, W., Kanbach, G., Manchester, R. N., Masnou, J. L., Mayer-Hasselwander, H. A., Ozel, M. E., Paul, J. A., Sacco, B., Scarsi, L., and Strong, A. W. (1983). Search for pulse γ-ray emission from ratio pulsars in the cos-b data. *Astronomy and Astrophysics*, 128(1):245–251.

Bulla, J., Chesneau, C., and Kachour, M. (2013). On the bivariate skellam distribution. *Communications in Statistics*. to appear.

Burkitt, A. N. (2006a). A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological Cybernetics*, 95(1):1–19.

Burkitt, A. N. (2006b). A review of the integrate-and-fire neuron model: Ii. inhomogeneous synaptic input and network properties. *Biological Cybernetics*, 95(2):97–112.

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33:325–340.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press, New York, NY, 1 edition.

Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE*, 6(1):e16104.

Calvagno, G., Ermani, M., Rinaldo, R., and Sartoretto, F. (2000). A multiresolution approach to spike detection in EEG. In *IEEE International Conference On Acoustics, Speech, And Signal Processing*.

Calvin, W. H. and Stevens, C. F. (1968). Synaptic noise and other sources of randomness in motoneuron interspike intervals. *Journal of Neurophysiology*, 31(4):574–587.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Prinston University Press, Princeton, NJ.

Caplan, J. B. and Glaholt, M. G. (2007). The roles of EEG oscillation in learning relational information. *Neuroimage*, 38(3):604–616.

Caplan, J. B., Madsen, J. R., Raghavachari, S., and Kahana, M. J. (2001). Distinct patterns of brain oscillations underline two basic parameters of human maze learning. *Journal of Neurophysiology*, 86:368–380.

Caplan, J. B., Madsen, J. R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E. L., and Kahana, M. J. (2003). Humanθ oscillations related to sensorimotor integration and spatial learning. *The Journal of Neuroscience*, 23(11):4726–4736.

Carandini, M., Horton, J. C., and Sincich, L. C. (2007). Thalamic filtering of retinal spike trains by postsynaptic summation. *Journal of Vision*, 7(14):1–11.

Carlson, N. R. (2007). *Physiology of Behavior*. Pearson, New York, NY, 9th edition.

Casagrande, V. and Ichida, J. (2011). Processing in the Lateral Geniculate Nucleus (lgn). In Alm, A., Kaufman, P.L. Levin, L., Nilsson, S., Ver Hoeve, J., and Wu, S., editors, *Adler's Physiology of the Eye*, pages 574–585. Elsevier, New York, NY, 11th edition.

Casella, G. and Berger, R. L. (2001). *Statistical Inference*. Duxbury Press, Canada, second edition.

Cateau, H. and Reyes, A. (2006). Relation between single neuron and population spiking statistics and effects on network activity. *Physical Review Letters*, 96(5):58101–1 – 58101–4.

Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of the Roya Statistical Society. Series B*, 14(2):220–228.

Chandra, R. and Optican, L. M. (1997). Detection, classification, and superposition resolution of action potentials in multiunit recordings by an online real time neural network. *IEEE Transactions in Biomedical Engineering*, 44(5):403–412.

Chechik, G. (2003). *An Information Theoretic Approach to the Study of Auditory Coding*. PhD thesis, The Hebrew University of Jerusalem. The Interdisciplinary Center for Neural Computation.

Chelaru, M. I. and Jog, M. S. (2005). Spike source localization with tetrodes. *Journal of Neuroscience M*, 142:305–315.

Chen, Z., Vijayan, S., Barbieri, R., Wilson, M. A., and Brown, E. N. (2009). Discrete- and continuous-time probabilistic models and algorithms for inferring neuronal up and down states. *Neural Computation*, 21(7):1797–1862.

Chernick, M. R. (1999). *Bootstrap Methods, A practitioner's guide*. John Wiley & Sons, New York, NY.

Churchland, P. and Sejnowski, T. (1994). *The Computational Brain*. Computational Neuroscience. A Bradford Book, fourth edition.

Cohen, M. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819.

Cooke, S. F. and Bliss, T. V. P. (2006). Plasticity in the human central nervous system. *Brain*, 129:159–173.

Criado, J. M., de la Fuente, A., Heredia, M., Riolobos, A. S., and Yajeya, J. (2008). Single-cell recordings: A method for investigating the brain's activation pattern during exercise. *Methods*, 45:262–270.

Cunningham, J. P., Yu, B., Shenoy, K., and Sahani, M. (2008). Inferring neural firing rates from spike trains using Gaussian processes. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20. MIT Press, Cambridge, MA.

Daley, D. and Vere-Jones, D. (2008). *An introduction to the theory of point process*. Springer, New York, NY, second edition.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM: Society for Industrial and Applied Mathematics, Philadelphia, PA.

Davison, A. C. and Hinkley, D. (2006). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, UK, 8th edition.

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, Cambridge, Massachusetts.

de Charms, C. R. and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, 381(6583):610–613.

de Charms, R. C. (2008). Applications of real-time fMRI. *Nature Reviews Neuroscience*, 9:720–729.

de Garis, H., Shuo, C., Goertzel, B., and Ruiting, L. (2010). A world survey of artificial brain projects, part i: Large-scale brain simulations. *Neurocomputing*, 74:3–29.

De Jager, O. C., Swanepoel, J. W. H., and Raubenheimer, B. C. (1989). A powerful test for weak periodic signals with unknown light curve shape in sparse data. *Astronomy and Astrophysics*, 221:180–190.

de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275:1805–1808.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B*, 60(2):333–350.

Destexhe, A. and Rudolph-Lilith, M. (2012). *Neuronal Noise*. Springer Series in Computational Neuroscience (Book 8). Springer.

Dimatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitiing with free-knot splines. *Biometrika*, 88(4):1055–1071.

Drewe, E. A. (1975). Go - no go learning after frontal lobe lesions in humans. *Cortex*, 11(1):8–16.

Eagleman, D. (2007). 10 unsolved mysteries of the brain. *Discover Magazine. Science, Technology and The Future*.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587.

Eden, U. T. (2008). Point process models for neural spike trains. Technical report, sfn.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., and Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, 16:971–998.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Eggermont, P. and LaRiccia, V. (2001). *Maximum Penalized Likelihood Estimation*. Springer, New York, NY.

Ekstrom, A. D., Caplan, J. B., Ho, E., Fried, I., and Kahana, M. J. (2005). Human hippocampal theta activity during virtual navigation. *Hi*, 15:881–889.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press, New York, NY.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, 338(6111):1202–1205.

Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9:292–303.

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):021905.

Fee, M. S., Mitra, P. P., and Kleinfeld, D. (1996). Variability of extracellular spike waveforms of cortical neurons. *Journal of Neurophysiology*, 76(6):3823–3833.

Ferguson, J. E. (2011). *Improving neural recording technology at the nanoscale*. PhD thesis, University of Minnesota.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(4):615–629.

Ferreira, M. A. R. and Lee, H. K. H. (2010). *Multiscale Modeling: A Bayesian Perspective*. Springer, New York, NY.

Fischer, Y., Wittner, L., Freund, T. F., and Gähwiler, B. H. (2002). Simultaneous activation of gamma and theta network oscillations in rat hippocampal slice cultures. *Journal of Physiology*, 539(3):857–868.

Fish, J., editor (2010). *Multiscale Methods: Bridging the Scales in Science and Engineering*. Oxford University Press, New York, NY.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., Pirice, C. J., Zeki, S., Ashburner, J., and Penny, W. (2004). *Human Brain Function*. Elsevier Academic Press, Chile, 2 edition.

Freeman, W. (1992). Predictions on neocortical dynamics derived from studies in paleocortex. In Basar, E. and Bullock, T., editors, *Induced Rhythms in the Brain*, pages 183–199. Birkhauser.

Gabbiani, F. and Koch, C. (1998). Principles of spike train analysis. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling: From Ions to Networks*, chapter 9. MIT Press, second edition.

Ganesan, M., Sumesh, E. P., and Vidhyalavanya, R. (2010). Multi-stage, multi-resolution method for automatic characterization of epileptic spikes in EEG. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 3(2):33–40.

Gavit, L., Baillet, S., Mangin, J., Pescatore, J., and Garnero, L. (2001). A multiresolution framework to MEG/EEG source imaging. *IEEE Transactions in Biomedical Engineering*, 48(10):1080–1087.

Gerstein, G. L. and Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal*, 4(1):41–68.

Gerstner, W. and Kistler, W. M. (2002). *Spiking Neuron Models: single neurons, populations, plasticity*. Cambridge University Press, Cambridge, UK.

Giuliodori, M. J. and Zuccolilli, G. (2004). Postsynaptic potential summation and action potential initiation: function following form. *Advances in Physiology Education*, 28(2):79–80.

Givens, G. and Hoeting, J. (2005). *Computational Statistics*. John Wiley & Sons, New York, NY.

Gnedin, A. V. (2007). The chain records. *Electronic Journal of Probability*, 12:767–786.

Goodale, M. A. and Milner, A. D. (2005). *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford University Press, New York, NY.

Gorban, A. N., Keql, B., Wunsch, D. C., and A., Z., editors (2007). *Principal Manifolds for Data Visualization and Dimension Reduction*. Springer, New York, NY.

Gourévitch, B. and Eggermont, J. J. (2007). A nonparametric approach for detection of bursts in spike trains. *Journal of Neuroscience Methods*, 160(2):349–358.

Gouwens, N. W., Zeberg, H., Tsumoto, K., Tateno, T., Aihara, K., and Robinson, H. P. C. (2010). Synchronization of firing in cortical fast-spiking interneurons at gamma frequencies: A phase-resetting analysis. *PLoS Computational Biology*, 6(9):1–13.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Academic Press, London, UK, seven edition.

Gray, C. M., Konig, P., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337.

Gray, C. M., Maldonado, P. E., Wilson, M., and McNaughton, B. (1995). Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *Journal of Neuroscience Methods*, 63:43–54.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Green, P. J. (1998). Penalized likelihood. In *Encyclopedia of Statistical Sciences, Update Volume 2*. Wiley-Interscience.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall/CRC, New York, NY.

Gregory, P. C. (1992). A new method for the detection of a periodic signal of unknown shape and period. *The Astrophysical Journal*, 398:146–168.

Griffiths, R. C., Milne, R. K., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21:238–255.

Grillenzoni, C. (2005). Non-parametric smoothing of spatio-temporal point processes. *Journal of Statistical Planning and Inference*, 128:61–78.

Grimmett, G. and D., S. (2001). *Probability and Random Processes*. Oxford University Press, New York, NY, third edition.

Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neuropathology*, 101(3):1126–1140.

Grun, S. and Rotter, S., editors (2010). *Analysis of Parallel Spike Trains*. Springer, New York, NY.

Hampela, D. and Lanskya, P. (2008). On the estimation of refractory period. *Journal of Neuroscience Methods*, 171:288–295.

Hanes, D. P., Thompson, K. G., and Schall, J. D. (1995). Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: Poisson spike train analysis. *Experimental Brain Research*, 103:85–96.

Harris, K. D., Henze, D. A., Hirase, H., Leinekugel, X., Dragoi, G., Czurko, A., and Buzsáki, G. (2002). Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature*, 417:738–741.

Hashorva, E. and Hüsler, J. (2005). Multiple maxima in multivariate samples. *Statistics and Probability Letters*, 75(1):11–17.

Haslinger, R., Pipa, G., and Brown, E. N. (2010). Discrete time rescaling theorem: determining goodness of fit for discrete time statistical models of neural spiking. *Neural Computation*, 22(10):2477–2506.

Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm? linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus*, 15:936–949.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, New York, NY.

Hebb, D. O. (2002). *The Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons, The Organization of Behavior: A Neuropsychological Theory.

Heikkinen, J. E., Rinne, R. I., Alahuhta, S. M., Lumme, J. A. J., Koivitso, M. E., Kirkinen, P. P., Sotaniemi, K. A., Nuutinen, L. S., and Jarvinen, P. A. (1985). Life support for 10 weeks with successful fetal outcome after fatal maternal brain damage. *British Medical Journal*, 290:1237–1238.

Heil, C. and Walnut, D. F. (2006). *Fundamental papers in wavelet theory*. Princeton University Press, Princeton, NJ.

Hernandez, E. and Guido, W. (1996). *A First Course on Wavelets*. CRC, New York, NY.

Hille, B. (2001). *Ion Channels of Excitable Membranes*. Sinauer Associates Inc., Sunderland, MA, 3 edition.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544.

Hughes, J. R. (1958). Post-tetanic potentiation. *Physiological Reviews*, 38:91–113.

Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Hwang, H.-K. and Tsai, T.-H. (2010). Multivariate records based on dominance. *Electronic Journal of Probability*, 15:1863–1892.

Indiradevi, K. P., Elias, E., Sathidevi, P. S., Nayak, S. D., and Radhakrishnan, K. (2008). A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Computers in Biology and Medicine*, 38(7):805–816.

Insel, N., Patron, L., Wagner, Z., Vega, J. N., and Barnes, C. A. (2010). Micro-scale coupling, macro-scale antagonism between excitation and inhibition during a decision-making task. In *Society for Neuroscience conference*.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society. Series A*, 100(3):415–416.

Izhikevich, E. M. (2000). Neural excitability, spiking and bursting. *International Journal of Bifurcation and Chaos*, 10(6):1171–1266.

Izhikevich, E. M. and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *PNAS*, 105:3593–3598.

Jensen, O. and Colgin, L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in cognitive sciences*, 11(7):267–269.

Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (2002). Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. *Cerebal Cortex*, 12:877–882.

Joeken, S., Schwegler, H., and Richter, C. P. (1997). Modeling stochastic spike train responses of neurons: an extended wiener series analysis of pigeon auditory nerve fibers. *Biological Cybernetics*, 76:153–162.

Jog, M. S., Connolly, C. I., Kubota, Y., Iyengar, D. R., Garrido, L., and Harlan, R. (2002). Tetrode technology: advances in implantable hardware, neuroimaging, and data analysis techniques. *Journal of Neuroscience Methods*, 117:141–152.

Johnson, D. and Ray, W. (2004). Optimal stimulus coding by neural populations using rate codes. *Journal of Computational Neuroscience*, 16:129–138.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 3(4):275–299.

Johnson, K. O. (1980). Sensory discrimination: neural processes preceding discrimination decision. *Journal of Neurophysiology*, 43:1793–1815.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley & Sons, New York, NY.

Joshi, S. W. and Patil, G. P. (1970). *Random Counts in Scientific Work*, volume 2, chapter A class of statistical models for multiple counts, pages 189–203. Pennsylvania State University Press.

Joshi, S. W. and Patil, G. P. (1971). Certain structural properties of the sum-symmetric power series distributions. *Sankhya-A*, 33(2):175–184.

Kalbfleisch, J. and Prentice, R. (1980). *The statistical analysis of failure time data*. John Wiley & Sons, New York, NY.

Kanai, R., Feilden, T., Firth, C., and Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, 21(8):677–680.

Kandel, E., Schwartz, J., and Jessel, T. (2000). *Principles of Neural Science*. McGraw-Hill Medical, New York, NY, 4 edition.

Karlis, D. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52(3):381–393.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15:255–265.

Kass, R. and Ventura, V. (2006). Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Computation*, 18:2583–2591.

Kass, R. E. and Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 2001(8):1713–1720.

Kass, R. E., Ventura, V., and Brown, E. (2005). Statistical issues in the analysis of neural data. *Journal of Neurophysiology*, 94:8–24.

Kass, R. E., Ventura, V., and Cai, C. (2003). Statistical smoothing of neuronal data. *Nework: Computation in Neural Systems*, 14:5–15.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to schwarz criterion. *Journal of American Statistical Association*, 90(431):928–934.

Kaufman, C. G., Ventura, V., and Kass, R. E. (2005). Spline-based non-parametric regression for periodic functions and its application to directional tuning of neurons. *Statistics in Medicine*, 24:2255–2265.

Kayser, C., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4):597–608.

Khatri, C. G. (1959). On certain properties of power-series distributions. *Biometrika*, 46(3-4):486–490.

Kim, B. and Basso, M. A. (2010). A probabilistic strategy for understanding action selection. *The Journal of Neuroscience*, 30(6):2340–2355.

Kim, S., Carmena, J. M., Nicoleis, M. A., and Principe, J. C. (2005). Multiresolution representations and data mining of neural spikes for brain-machine interfaces. In *2nd International IEEE EMBS Conference on Neural Engineering*.

Kimura, D. (1977). Acquisition of motor skill after left hemisphere damage. *Brain*, 100:527–542.

Kistler, W. and De Zeeuw, C. (2002). Dynamical working memory and timed responses: the role of reverberating loops in the olivo-cerebellar system. *Neural Computation*, 14(11):2597–2626.

Kitazawa, S. and Wolpert, D. M. (2005). Rhythmisity, randomness and synchrony in climbing fiber signals. *Trends in Neurosciences*, 28(11):611–619.

Knapen, T., Brascamp, J., Pearson, J., van Ee, R., and Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *Journal of Neuroscience*, 31(28):10293–10301.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker Inc., New York, NY.

Kolaczyk, E. and Nowak, R. (2005). Multiscale generalised linear models for nonparametric function estimation. *Biometrika*, 92(1):119–133.

Kolaczyk, E. D. (1999). Bayesian multiscale models for Poisson processes. *Journal of the American Statistical Association*, 94(447):920–933.

Kolaczyk, E. D. and Nowak, R. D. (2003). Nonlinear estimation and classification. In Denison, D. D., Hansen, M. H., Holmes, C. C., Mallick, B., and Yu, B., editors, *Lecture Notes in Statistics*. Springer.

Kolaczyk, E. D. and Nowak, R. D. (2004). Multiscale likelihood analysis and complexity penalized estimation. *The Annals of Statistics*, 32:500–527.

Kolb, B. and Whishaw, I. (2008). *Fundamentals of Human Neuropsychology*. Worth Publishers, New York, NY.

Komoya, S. and Kass, R. E. (2008). Spike train probability models for stimulus-driven leaky integrate-and-fire neurons. *Neural Computation*, 20(7):1776–95.

Köppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. *The Journal of Neuroscience*, 17(9):3312–3321.

Kostal, L., Lansky, P., and Rospars, J. (2007). Neuronal coding and spiking randomness. *European Journal of Neuroscience*, 26:2693–2701.

Koyama, S., Eden, U., Brown, E. N., and Kass, R. E. (2010). Bayesian decoding of neural spike trains. *Annals of the Institute of Statistical Mathematics (AISM)*, 62:37–59.

Lagarias, J., Reeds, J., Wright, M., and Wright, P. (1998). Lagarias jc, reeds ja, wright mh, wright pe. *SIAM Journal of Optimization*, 9(1):112–147.

Laing, C. and Lord, G. J., editors (2009). *Stochastic Methods in Neuroscience*. Oxford University Press, New York, NY.

Lakshminarayana, J., Pandit, S. N. N., and Rao, K. S. (1999). On a bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 28(2):267–276.

Lange, G. D. and Hartline, P. H. (1979). Fourier analysis of spike train data. *Biological Cybernetics*, 34:31–34.

Laubach, M. (2004). Wavelet-based processing of neural spike trains prior to discriminant analysis. *Journal of Neuroscience Methods*, 134:159–168.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, 23(4/5):727–738.

Lee, J., Kim, H. R., and Lee, C. (2010). Trial-to-trial variability of spike response of V1 and saccadic response time. *Journal of Neurophysiology*, 104:2556–2572.

Lefkimmiatis, S., Maragos, P., and Papandreou, G. (2009). Bayesian inference on multiscale models for Poisson intensity estimation: Applications to photon-limited image denoising. *IEEE Transactions on Image Processing*, 18(8):1724–1741.

Legatt, A. D., Arezzo, J., and Vaughan, H. G. (1980). Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. *Journal of Neuroscience Methods*, 2(2):203–217.

Lehky, S. (2004). Bayesian estimation of stimulus responses in Poisson spike trains. *Neural Computation*, 16(7):1325–1343.

Lehky, S. R. (2010). Decoding Poisson spike trains by Gaussian filtering. *Neural Computation*, 22(5):1245–71.

Lehmann, E. L. (1983). *Theory of point estimation*. John Wiley & Sons, New York, NY.

Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computational Neural System*, 9(4):R53–R78.

Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *Journal of Sound and Vibration*, 12(3):252–275.

Lewis, P. A. W., editor (1972). *Stochastic Point Processes*, chapter Recent Results in the Statistical Analysis of Univariate Point Processes, pages 1–54. John Wiley & Sons.

Li, C. T., Poo, M., and Dan, Y. (2009). Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324:643–646.

Lipton, B. H. (2005). *The Biology of Belief*. Hay House Inc., Carlsbad, California.

Lisman, J. E. and Idiart, M. A. P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, 267(5203):1512–1515.

Llinás, R. R. and Steriade, M. (2006). Bursting of thalamic neurons and states of vigilance. *Journal of Neuroohysiology*, 95:3297–3308.

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and clinical Neurophysiology*, 79:81–93.

López-Muñoz, F., Boya, J., and Alamo, C. (2006). Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel prize award to santiago ramón y cajal. *Brain Research Bulletin*, 70:391–405.

Lytton, W. W. (2002). *From Computer to Brain: Foundations of Computational Neuroscience*. Springer, New York, NY.

Magistretti, P. J., Pellerin, L., and Martin, J. L. (1995). Brain energy metabolism: An integrated cellular perspective. In Bloom, F. E. and Kupfer, D. J., editors, *Psychopharmacology: the Fourth Generation of Progress*, pages 657–670. Raven Press, New York, NY.

Maimon, G. and Assad, J. A. (2009). Beyond Poisson: Increased spike-time regularity across primate parietal cortex. *Neuron*, 62(3):426–440.

Majumdar, K. K. (2007). A structural and a functional aspect of stable information processing by the brain. *Cognitive Neurodynamics*, 1:295–303.

Malhotra, S., Cross, R., and van der Meer, M. (2012). Theta phase precession beyond the hippocampus. *Reviews in the Neurosciences*, 23(1):39–65.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mandelbrot, B. B. (1999). *Multifractals and 1/f Noise: Wild Self-Affinity in Physics*. Springer, New York, NY.

Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, 7:153–160.

Marshall, A. W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of American Statistical Association*, 80:332–338.

Martin, R. (1998). *Neuroscience Methods: A Guide for Advanced Students*. CRC Press, Amsterdam, Netherlands.

Martin, T. and Urbanek, S. (2008). *Interactive Graphics for Data Analysis: Principles and Examples*. Chapman & Hall, Boca Raton, FL.

Massey, P. V. and Bashir, Z. I. (2007). Long-term depression: multiple forms and implications for brain function. *Trends in Neurosciences*, 30(4):176–184.

Maurer, A. P. and McNaughton, B. L. (2007). Network and intrinsic cellular mechanisms underlying theta phase precession of hippocampal neurons. *Trends in Neurosciences*, 30(7):325–333.

McKendrick, A. G. (1925). Applications of mathematics to medical problems. In *Proceedings of the Edinburgh Mathematical Society*, volume 44, pages 98–130.

McNaughton, B. L., O'Keefe, J., and Barnes, C. A. (1983). The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *Journal of Neuroscience Methods*, 8(4):391–397.

Mehta, M. R., Lee, A. K., and Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, 417(13):741–746.

Meinshausen, N., Bickel, P., and Rice, J. (2009). Efficient blind search: optimal power of detection under computational cost constrains. *The Annals of Applies Statistics*, 3(1):38–60.

Meng, L., Kramer, M. A., and Eden, U. T. (2011). A sequential monte carlo approach to estimate biophysical neural models from spikes. *Journal of Neural Engineering*, 8:065006.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge, UK.

Miller, R. (1991). *Cortico-Hippocampal Interplay and the Representation of Contexts in the Brain*. Springer, New York, NY.

Milner, B. (2003). Visual recognition and recall after right temporal-lobe excision in man. *Epilepsy and Behavior*, 4(6):799–812.

Milpass, K. (2012). Movement disorders: Motor changes in sleep disorder signal prodromal parkinsonism. *Nature Reviews Neuroscience*, doi:10.1038/nrneurol.2012.104.

M'Kendrick, A. G. (1925). Applications of mathematics to medical problems. In *Proceedings of the Edinburgh Mathematical Society*, volume 44, pages 98–130.

Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K., and Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current Biology*, 18(5):375–380.

Montgomery, S. M., Sirota, A., and Buzsáki, G. (2008). Theta and gamma coordination of hippocampal networks during walking and rem sleep. *Journal of Neuroscience*, 28(26):6731–6741.

Moore, G. P., Perkel, D., and Segundo, J. P. (1966). Statistical analysis and functional interpretation of neural spike data. *Annual Review of Physiology*, 28:493–522.

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences*, 11(11):489–497.

Napadow, V., Dhond, R., Conti, G., Makris, N., Brown, E. N., and Barbieri, R. (2008). Brain correlations of automatic modulation: Combining heart rate variability with fMRI. *Neuroimage*, 42(1):169–177.

Natarajan, R. (2003). Uber-Claws: Unsupervised pattern classification for multi-unit extracellular neural burst extraction. Master's thesis, New Jersy Institute of Technology.

Nawrot, M. P., Boucsein, C., Rodriguez, M. V., Riehle, A., Aertsen, A., and Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods*, 169(2):374–390.

Neason, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer, New York, NY.

Nedungadi, A. G., Rangarajan, G., Jain, N., and Ding, M. (2009). Analyzing multiple spike trains with nonparametric granger causality. *Journal of Computational Neuroscience*, 27(1):55–64.

Nelder, J. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, 7:308–313.

Nelson, M. E. (2002). Multiscale spike train variability in primary electrosensory afferents. *Journal of Physiology - Paris*, 96:507–516.

Nicholls, J. G., Martin, A. R., Fuchs, P. A., Brown, D. A., Diamond, M. E., and Weisblat, D. A. (2012). *From Neuron to Brain*. Sinauer Associates, Inc., Sunderland, MA, 5 edition.

Nicolelis, M. A. L., editor (2001). *Advances in Neural Population Coding*. Elsevier Science, London, UK.

Noack, A. (1950). A class of random variables with discrete distributions. *The Annals of Mathematical Statistics*, 21(1):127–132.

Nowak, R. D. and Kolaczyk, E. D. (2000). A statistical multiscale framework for Poisson inverse problems. *IEEE Transactions on Information Theory*, 46:1811–1825.

Nuñez, A., Amzica, F., and Steriade, M. (1992). Intrinsic and synaptically generated delta (1-4hz) rhythms in dorsal lateral geniculate neurons and their modulation by light-induced fast (30-70hz) events. *Neuroscience*, 51(2):269–284.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(12):171–175.

O'Keefe, J. and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330.

Øksendal, B. (2010). *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York, NY.

Omi, T. and Shinomoto, S. (2011). Optimizing time histograms for non-Poissonian spike trains. *Neural Computation*, 23(12):3125–44.

Onken, A., Gr unew alder, S., Munk, M. H. J., and Obermayer, K. (2009). Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation. *PLOS Computational Biology*, 5:1–13.

Oulfsen, M., Whittington, M., Camperi, M., and Kopell, N. (2003). New roles for the gamma rhythm: population tuning and preprocessing for the beta rhythm. *Journal of Computational Neuroscience*, 14(14):33–54.

Paddock, S. M., Ruggeri, F., Lavine, M. L., and West, M. (2003). Randomized Polya tree models for nonparametric Bayesian inference. *Statistica Sinica*, 13:443–460.

Palm, G., Aertsen, A. M. H. J., and Gerstein, G. L. (1988). On the significance of correlations among neuronal spike trains. *Biological Cybernetics*, 59(1):1–11.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15:243–262.

Paninski, L. (2010). Fast Kalman filtering on quasilinear dendritic trees. *Journal of Computational Neuroscience*, 28(2):211–228.

Paninski, L., Brown, E. N., Iyengar, S., and Kass, R. E. (2008). Statistical models of spike trains. In Laing, C. and Lord, G., editors, *Stochastic Methods in Neuroscience*. Oxford University Press.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. In Cisek, P., Drew, T., and Kalaska, J., editors, *Computational Neuroscience*, volume 165 of *Progress in Brain Research*, pages 493–507. Elsevier Science.

Paninski, L., Pillow, J. W., and Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Computation*, 16(12):2533–61.

Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M., and Diamond, M. E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29(3):769–777.

Park, I., Seth, S., Paiva, A. R. C., Li, L., and Principe, J. C. (2013). Kernel methods on spike train space for neuroscience: a tutorial. *IEEE Signal Processing Magazine*, pages 149–160.

Park, I., Seth, S., Rao, M., and Principe, J. (2012). Strictly positive-definite spike train kernels for point-process divergences. *Neural Computation*, 24(8):2223–50.

Patil, G. P. (1962). Certain properties of the generalized power series distribution. *Annals of the Institute of Statistical Mathematics*, 14(1):179–182.

Patil, G. P. (1968). On sampling with replacement from populations with multiple characters. *Sankhya-B*, 30(3/4):354–366.

Pauler, D. K. (1998). The Schwarz criterion and related methods for normal linear models. *Biometrika*, 85(1):13–27.

Paulin, M. G. (1992). Digital filters for firing rate estimation. *Biological Cybernetics*, 66:525–531.

Pawlas, Z., Klebanov, L. B., Benes, V., Prokesová, M., Popelár, J., and Lánský, P. (2010). First-spike latency in the presence of spontaneous activity. *Neural Computation*, 22(7):1675–1697.

Pearce, J. M. (2001). Emil Heinrich Du Bois-Reymond (1818-96). *Journal of Neurology Neurosurgery and Psychiatry*, 71(5):620.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neural spike trains and stochastic point processes: I. the single spike train. *Biophysical Journal*, 7(4):391–418.

Phillips, R. (2001). *Crystals, Defects and Microstructures*. Cambridge University Press, Cambridge, UK.

Piekema, C., Kessels, R. P. C., Mars, R. B., Petersson, K. M., and Fernandez, G. (2006). The right hippocampus participates in short-term memory maintenance of object-location associations. *Neuroimage*, 33(1):374–382.

Pillow, J. (2007). Likelihood-based approaches to modeling the neural code. In Doya, K., Ishi, S., Pouget, A., and Rao, R. P. N., editors, *Bayesian Brain: Probabilistic approaches to neural coding*. MIT Press.

Pillow, J., Ahmadian, Y., and Paninski, L. (2010). Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Computation*, 23(1):1–45.

Pillow, J. W. (2009). Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding model. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*. MIT Press.

Pillow, J. W. and Scott, J. G. (2012). Fully bayesian inference for neural models with negative-binomial spiking. *Advances in Neural Information Processing Systems (NIPS)*, 25. to appear.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., and Chichilnisky, E. (2008). Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature*, 454:995–1001.

Poppe, C., Martens, G., De Bruyne, S., Lambert, P., and Van de Walle, R. (2008). Robust spatio-temporal multimodal background subtraction for video surveillance. *Optical Engineering*, 47(10):1–13.

Pouzat, C. (2008). Techniques for spike sorting. arXiv:q-bio/0405012.

Pouzat, C., Chaffiol, A., and Gu, C. Static and dynamic models for spike train analysis. to appear in Journal of Computational Neuroscience.

Protassov, R. and Van Dyk, D. A. (2002). Statistics, handle with care: detecting multiple model components with the likelihood ratio test. *The Astrophysical Journal*, 571:545–559.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A. S., McNamara, J. O., and White, L. E. (2008a). *Neuroscience*. Sinauer Associates, Inc., Sunderland, MA, 4 edition.

Purves, D., Brannon, E. M., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., and Woldorff, M. G. (2008b). *Principles of Cognitive Neuroscience*. Sinauer Associates, Inc., Sunderland, MA.

Purvis, L. K. and Butera, R. J. (2005). Inonic current model of a hypoglossal motoneuron. *Journal of Neurophysiology*, 93(2):723–733.

Puzzo, I., Cooper, N. R., Vetter, P., and Russo, R. (2010). EEG activation differences in the pre-motor cortex and supplementary motor area between normal individuals with high and low traits of autism. *Brain Research*, 1342:104–110.

Ramezan, R., Marriott, P., and Chenouri, S. (2010). An inhibitory-excitatory approach for the analysis of the neural spike trains. In *The Joint Statistical Meeting (JSM)*, Vancouver, Canada. American Statistical Association.

Ramezan, R., Marriott, P., and Chenouri, S. (2012). Skellam process and its application in modelling neural spike trains. In *Annual meeting of the Society for Neuroscience*, New Orleans, USA.

Ramezan, R., Marriott, P., and Chenouri, S. (2014). Multiscale analysis of neural spike trains. *Statistics in Medicine*, 33(2):238–256.

Ramón y Cajal, S. (1954). *Neuron theory or reticular theory?: Objective evidence of the anatomical unity of nerve cells*. Consejo Superior de Investigaciones Cientificas.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data Analysis*. Springer, New York, NY, second edition.

Reich, D. S., Mechler, F., and Victor, J. D. (2001). Independent and redundant information in nearby cortical neurons. *Science*, 294(5551):2566–2568.

Reid, N. (2010). Likelihood inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):517–525.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.

Renshaw, A., Forbes, A., and Morison, B. R. (1939). Activity of isocortex and hippocampus: electrical studies with micro-electrodes. *Journal of Neurophysiology*, 3(1):74–105.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. The MIT Press, London, England.

Rigas, A. G. and Tsitsis, D. S. (1996). Spectral analysis techniques of stationary point processes: Extensions and applications to neurophysiological problems. *Computers & Mathematics with Applications*, 32(11):93–99.

Robinson, P. A. and Rennie, C. (2007). Quantitative modeling of multiscale neural activity. In Bender, A., editor, *Complexity and Nonlinear Dynamics*, volume 6417 of *Proceedings of the SPIE*, page 64170F.

Rullen, R. V. and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6):1255–1283.

Salimpour, Y., Soltanian-Zadeh, H., Salehi, S., Emadi, N., and Abouzari, M. (2011). Neuronal spike train analysis in likelihood space. *PLOS ONE*, 6(6):e21256.

Sanchez, J. C., Mareci, T. H., Norman, W. M., Principe, J. C., Ditto, W. L., and Carney, P. R. (2006). Evolving into epilepsy: Multiscale electrophysiological analysis and imaging in an animal model. *Experimental Neurology*, pages 31–47.

Sanchez, J. C. and Principe, J. C. (2007). *Brain-Machine Interface Engineering*. Morgan and Claypool Publishers.

Sanger, T. D. (2002). Decoding neural spike trains: Calculating the probability that a spike train and an external signal are related. *Journal of Neurophysiology*, 87:1659–1663.

Sarma, S. V., Eden, U. T., Cheng, M. L., Williams, Z. M., Hu, R., Eskandar, E., and Brown, E. N. (2010). Using point process models to compare neural spiking activity in the subthalamic nucleus of parkinson's patients and a healthy primate. *IEEE Transactions in Biomedical Engineering*, 57(6):1297–1307.

Schmid, C. M., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., Ye, F. Q., and Leopold, D. A. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature*, 466:373–377.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(20):1007–1012.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Schwarzenberg-Czerny, A. (1999). Optimum period search: quantitative analysis. *The Astrophysical Journal*, 516:315–323.

Shadlen, M. N. and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579.

Shao, N. and Lii, K. (2011). Modelling non-homogeneous Poisson processes with almost periodic intensity functions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 73:99–122.

Sherman-Gold, R. (1993). *The Axon Guide*. Axon Instruments, Inc., USA.

Sherrington, C. (1906). *The Integrative Action of the Nervous System*. Cambridge University Press, Cambridge, UK.

Sherrington, C. S. (1926). *The Integrative Action of the Nervous System*. Yale University Press, New Haven, CT.

Shimazaki, H. and Shinomoto, S. (2006). A recipe for optimizing a time-histogram. In *Advances in Neural Information Processing Systems*, volume 19, pages 1289–1296, Cambridge, MA. MIT Press.

Shin, H., Aggarwal, V., Acharya, S., Schieber, M. H., and Thakor, N. V. (2010). Neural decoding of finger movements using Skellam-based maximum likelihood decoding. *IEEE Transactions in Biomedical Engineering*, 57(3):754–760.

Shinomoto, S., Miura, K., and Koyama, S. (2005). A measure of local variation of inter-spike intervals. *BioSystems*, 79:67–72.

Simoes, C., Jensen, O., Parkkonen, L., and Hari, R. (2003). Phase locking between human primary and secondary somatosensory cortices. *PNAS*, 100(5):2691–2694.

Simonoff, J. (1996). *Smoothing Methods in Statistics*. Springer, New York, NY.

Sincich, L. C., Adams, D. L., Economides, J. R., and Horton, J. C. (2007). Transmission of spike trains at the retinogeniculate synapse. *The Journal of Neuroscience*, 27:2683–2692.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24:49–65.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society. Series A*, 109(3):296.

Slavin, G. S. and Bluemake, D. A. (2005). Spatial and temporal resolution in cardiovascular mr imaging: Review and recommendations. *Radiology*, 234:330–338.

Smith, A. C., Scalon, J. D., Wirth, S., Yanike, M., Suzuki, W. A., and Brown, E. N. (2010). State-space algorithms for estimating spike rate functions. *Computational Intelligence and Neuroscience*, 2010:1–14.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience*, 13(1):334–350.

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *PHYSICAL REVIEW E,*, 64:1–11.

Srinivasan, L., Eden, U. T., Mitter, S. K., and Brown, E. N. (2007). General-purpose filter design for neural prosthetic devices. *Journal of Neurophysiology*, 98(4):2456–2475.

Stapleton, J. R., Lavine, M. L., Wolpert, R. L., Nicolelis, M. A. L., and Simon, S. A. (2006). Rapid taste responses in the gustatory cortex during licking. *The Journal of Neuroscience*, 26(15):4126–4138.

Stein, R., Gossen, E. R., and Jones, K. E. (2005). Neuronal variability: noise or part of the signal? *Nature Reviews Neuroscience*, 6:389–397.

Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal*, 5(2):173–194.

Stein, R. B. (1967). The information capacity of nerve cells using a frequency code. *Biophysical Journal*, 7:797–826.

Stevens, M. C., Pearlson, G. D., and Kiehl, K. A. (2007). An fMRI auditory oddball study of combined-subtype attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 164(11):1737–1749.

Strumwasser, F. (1958). Long-term recording from single neurons in brain of unrestrained mammals. *Science*, 127(3296):469–470.

Suhail, Y. and Oweiss, K. (2004). Multiresolution Bayesian detection of multiunit extracellular spike waveforms in multichannel neuronal recordings. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

Takahashi, N., Sasaki, T., Usami, A., Matsuki, N., and Ikegaya, Y. (2007). Watching neuronal circuit dynamics through functional multineuron calcium imaging (fMCI). *Neuroscience Research*, 58:219–225.

Taketani, M. and Baudry, M., editors (2006). *Advances in Network Electrophysiology Using Multi-Electrode Arrays*. Springer, New York, NY.

Taniguchi, K. and Ogawa, H. (1987). Spectral analysis of instantaneous frequency responses to sinusoidal stimulation in cutaneous mechanoreceptor afferent units of frogs. *The Japanese Journal of Physiology*, 37(6):1031–1049.

Tenenbaum, M. and Pllard, H. (1963). *Ordinary Differential Equations*. Harper & Row.

Theunissen, F. and Miller, J. P. (1995). Temporal encoding in nervous systems: A rigorous definition. *Journal of Computational Neuroscience*, 2(2):149–162.

Thompson, C. and Duncan, R. (1993). Neutron star dynamos and the origins of pulsar magnetism. *The Astrophysical Journal*, 408:194–217.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520–522.

Tokdar, S., Xi, P., Kelly, R., and Kass, R. (2010). Detection of bursts in extracellular spike trains using hidden semi-markov point process models. *Journal of Computational Neuroscience*, 29(1-2):203–212.

Tort, A., Komorowski, R., Manns, J., Kopell, N., and Eichenbaum, H. (2009). Theta-gamma coupling increases during the learning of item-context associations. *PNAS*, 106(49):20942–47.

Tort, A., Kramer, M., Thorn, C., Gibson, D., Kubota, Y., Graybiel, A., and Kopell, N. (2008). Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a t-maze task. *PNAS*, 105:20517–22.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93:1074–1089.

Truccolo, W., Friehs, H. M., Donoghue, J. P., and Hochberg, L. R. (2008). Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. *Journal of Neuroscience*, 28(5):1163–1178.

Truccolo, W., Hochberg, L. R., and Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience*, 13(1):105–113.

Unwin, A., Martin, T., and Hofmann, H. (2006). *Graphics of Large Datasets: Visualizing a Million*. Springer, New York, NY.

Uzuntarla, M., Ozer, M., and Guo, D. Q. (2012). Controlling the first-spike latency response of a single neuron via unreliable synaptic transmission. *The European Physical Journal B*, 85(282):1–8.

van der Meer, M. A. and Redish, D. A. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *Frontiers in Integrative Neuroscience*, 3(1):1–15.

van Hemmen, J. L. and Sejnowski, T. J., editors (2006). *23 problems in systems neuroscience*. Oxford University Press, New York, NY.

van Vreeswijk, C. (2010). Stochastic models of spike trains. In Grün, S. and Rotter, S., editors, *Analysis of Parallel Spike Trains*. Springer.

Vanables, W. and Ripley, B. (1999). *Modern Applied Statistics with S-plus*. Springer, New York, NY, 3rd edition.

Varela, F., Lachaux, J., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2:229–239.

Ventura, V. (2004). Testing for and estimating latency effects for Poisson and non-Poisson spike trains. *Neural Computation*, 16(11):2323–49.

Vere-Jones, D. (1982). On the estimation of frequency in point-process data. *Journal of Applied Probability*, 19:383–394.

Vere-Jones, D. (1992). Statistical methods for the description and display of earthquake catalogs. In Walden, A. and Guttorp, P., editors, *Statistics in the Environmental and Earth Sciences*, pages 220–246. Arnold, London, England.

Victor, J. D. (1999). Temporal aspects of neural coding in the retina and lateral geniculate. *Network: Computational Neural System*, 10:R1–R66.

Victor, J. D. (2006). Approaches to information-theoretic analysis of neural activity. *Biological Theory*, 1(3):302–316.

Walker, K. M. M., Bizley, J. K., King, A. J., and Schnupp, J. W. H. (2011). Multiplexed and robust representations of sound features in auditory cortex. *The Journal of Neuroscience*, 31(41):14565–14576.

Walker, P. (1999). *Chambers Dictionary of Science and Technology*. Chambers, London, UK.

Walnut, D. F. (2002). *An Introduction to Wavelet Analysis*. Birkhäuser, Boston, MA.

Webster, D. B. (1999). *Neuroscience of Communication*. Singular Publication Group, Inc., 2 edition.

Weiss, R. D., Mirin, S. M., and Bartel, R. (2002). *Cocaine*. American Psychiatric Pub, Arlington, VA, 2 edition.

Whishaw, I. Q. and Vanderwolf, C. H. (1973). Hippocampal EEG and behavior: Changes in amplitude and frequency of RSA (theta rhythm) associated with spontaneous and learned movement patterns in rats and cats. *Behavioral Biology*, 8:461–484.

Whitten, T. A., Hughes, A. M., Dickson, C. T., and Caplan, J. B. (2011). A better oscillation detection method robustly extracts EEG rhythms across brain state changes: The human alpha rhythm as a test case. *Neuroimage*, 54:860–874.

Wiesel, T. and Hubel, D. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology - London*, 148:574–591.

Willett, R. and Nowak, R. (2002). Multiresolution nonparametric intensity and density estimation. In *IEEE International Conference on Acoustics , Speech, and Signal Processing*, volume 2, pages 1493–1496.

Wimmer, G. and Altmann, G. (1996). The multiple Poisson distribution, its characteristics and a variety of forms. *Biometrical Journal*, 38(8):995–1011.

Windhorst, U. and Johansson, H., editors (1999). *Modern Techniques in Neuroscience Research*. Springer, New York, NY.

Winocur, G., Moscovitch, M., Rosenbaum, R. S., and Sekeres, M. (2010). An investigation of the effects of hippocampal lesions in rats on pre- and postoperatively acquired spatial memory in a complex environment. *Hippocampus*, 20(12):1350–1365.

Wise, K. D. (2007). Integrated sensors, MEMS, and microsystems: Reflections on a fantastic voyage. *Sensors and Actuators, A: Physical*, 136(1):39–50.

Wu, W., Gao, Y., E., B., Donoghue, J., and Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Computation*, 18(1):80–118.

Young, M. D., Manchester, R. N., and Johnston, S. (1999). A radio pulsar with an 8.5-second period that challenges emission models. *Nature*, 400:848–849.

Zigmond, M. J., Bloom, F. E., Landis, S. C., Roberts, J. L., and Squire, L. R. (1999). *Fundamental Neuroscience*. Academic Press, London, UK.