Understanding Complex Systems

Founding Editor: J.A. Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition - typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels - and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuroand cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Leandro Pardo,
Narayanaswamy Balakrishnan,
and María Ángeles Gil (Eds.)

Modern Mathematical Tools and Techniques in Capturing Complexity

Springer

Editors

Leandro Pardo
Departamento de Estadística e I. O.
Facultad de Matemáticas
Universidad Complutense de Madrid
28040-Madrid Spain
E-mail: lpardo@mat.ucm.es
http://www.ucm.es/dir/gi001.htm

María Ángeles Gil
Departamento de Estadística e I. O. y D.M.
Universidad de Oviedo
C/ Calvo Sotelo, s/n 33071 Oviedo
Spain
E-mail: magil@uniovi.es
http://bellman.ciencias.uniovi.es/SMIRE

Narayanaswamy Balakrishnan
Department of
Mathematics and Statistics
McMaster University Hamilton,
Ontario Canada L8S 4K1
E-mail: bala@univmail.cis.mcmaster.ca
http://www.math.mcmaster.ca/bala/

ISBN 978-3-642-20852-2 e-ISBN 978-3-642-20853-9

DOI 10.1007/978-3-642-20853-9

Understanding Complex Systems ISSN 1860-0832

Library of Congress Control Number: 2011928064

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

This book is prepared as a tribute to
Professor María Luisa Menéndez
who was not only an exceptional researcher and teacher,
but was also a person with many notable skills and
and outstanding human qualities

Autocorrelation Measures and Independence Tests in Spike Trains

Aldana González-Montoro[1], Ricardo Cao[1], Nelson Espinosa[2], Jorge Mariño[2], and Javier Cudeiro[2]

[1] Department of Mathematics, Universidade da Coruña, Facultad de Informática, Campus Elviña, 15071 A Coruña, Spain
agonzalezmo@udc.es, rcao@udc.es
[2] Neuroscience and Motor Control Group (NEUROcom), Department of Medicine, Universidade da Coruña, Campus de Oza, 15006 A Coruña, Spain
nespinosa@udc.es, xurxo@udc.es, jcud@udc.es

Summary. In the nervous system, neurons convey information by means of electric pulses called action potentials or spikes. The information is encoded in sequences of these pulses, called spike trains. In neurophysiological experiments, spike trains are recorded and analyzed statistically. Time intervals between action potentials is the key feature of spike trains.

One of the statistical techniques available for studying spike trains is the autocorrelation. In the neuroscientific literature the term autocorrelation is used to denote frequency histograms of time intervals between every pair of the spikes generated by a single neuron for a period of time. The autocorrelation function is very useful for characterizing spike trains. The shape of the autocorrelation indicates the nature of the dependence among time intervals between consecutive spikes for a specified time window. In this work we propose two statistics to test the hypothesis of independence between time intervals. The bootstrap method is used to calibrate the null distribution of these tests. The tests are applied to real spike trains recorded from the primary visual cortex of anesthetized cats, both during spontaneous activity and after electric stimulation-induced activity.

1 Introduction

Neurons are, together with glial cells, the basic structural and functional units of the nervous system. One of the most important characteristics of neuronal cells is their ability to propagate big quantities of information, at fairly fast speeds throughout neural networks. The information is conveyed along the nervous system in the form of electrical signals called action potentials (AP) or spikes traveling through cable-like cellular extensions called *axons*.

Neurons are highly specific cells whose structure and physiology make possible the generation and transmission of AP. Roughly speaking, when the

input signals that continuously arrive to the initial segment of the axon reach a certain threshold, an abrupt change in the cells electrical membrane potential takes place, giving rise to an AP. Those electrical pulses travel along the membrane of the axon, and their trains are used as a binary code for the transfer of information between cells. Since AP are sharp potential changes, they are relatively easy to record. Researchers do so by placing electrodes close to, or inside, the neurons. In mammals APs have an approximate amplitude of 100 mV and a duration of 1 ms. The shape of the spikes remains practically unchanged while they travel through the axon, so the information they carry must be coded not in each AP but in a sequence of them.

A sequence of APs is called a *spike train*. Spike trains are very important because a great part of the information transmitted by neurons is coded in them. Spike trains are the object of study of the present work.

The global brain activity, i.e., the level of arousal and attentiveness, is modulated by the so called *activating ascending pathways*, constituted by neuronal nuclei situated in the brainstem (*bs*) and the basal forebrain (*bf*). This modulation is responsible of the main changes of activity that take place during the sleep-wake cycle. During the slow sleep state, neural activity is highly synchronized, which is reflected in the electroencephalogram (EEG) by waves of more amplitude and less frequency than in the wake state. The anesthetized state is very similar to and mimics the slow sleep state that occurs under physiological conditions, characterized by low frequency oscillatory activity. The arousal state, i.e., the transition from sleep to awake, can be induced by stimulating either the *bs* or the *bf*. In this work, simultaneous neuronal recordings made in the primary visual cortex of anesthetized cats were used to study the effects induced by the electrical stimulation of both pathways. Discussion on spike trains and modeling of neural systems can be found in [2] and a complete description of the state of the art on statistical analysis tools for multiple spike train data can be found in [1].

The following sections of the paper are organized as follows. Section 2 describes the experimental data. Two correlation measures for spike trains are introduced in Section 3. Section 4 deals with the statistics proposed to test independence between interspike intervals and include the results obtained using these methods. Finally, Section 5 contains the main conclusions.

2 Experimental Data

The experiments which yielded the data were performed in anesthetized cats. An eight-point multielectrode was introduced in the primary visual cortex in order to make simultaneous extracellular recordings of several neurons. Concurrently, an EEG was made and two other electrodes were introduced for electrical stimulation at *bs* and *bf*. The stimuli were electric pulses of 2 s of duration (trains of 0.05 s micro-pulses at a frequency of 50 Hz delivered for 2 s) which were applied differentially in the areas under study according to the following protocol. First, a group of neurons was identified and isolated

using the multielectrode, and their spontaneous activity was recorded for 2 minutes. Then, electrical stimulation was delivered either to *bs* or *bf* (the sequence of stimulation of those areas was randomized) for two seconds and, after another period of time (8 min), enough for the neurons to return to their spontaneous activity, the other region (*bs* or *bf*) was stimulated following the same procedure. Finally, the recording continued for another amount of time that allowed the neurons to return to spontaneous activity again.

Each of the eight electrodes of the multielectrode device may lay close enough to no, one or more than one neurons. Hence, under favorable experimental conditions, it is possible to record more than eight neurons simultaneously. In this work we deal with the simultaneous recording of seven neurons. We used three different recordings (called trials) for each stimulus and each neuron.

In our context, one trial is the recording of one neuron during the spontaneous activity followed by the application of the stimulus (either *bs* or *bf*) and a final time period for recovery. Each trial had a duration of around 600 seconds and stimuli were applied approximately after 120 seconds of the beginning of the recording, but not all of them occurred at the same instant. Original data are presented as time instants that indicate spike events (with a millisecond precision). The neurons are denoted N1, N3a, N3b, N4a, N4b, N5 and N7.

Since the aim of the neurophysiological work was to characterize the differential effects of the stimuli on neuronal activity, the recording of each trial was separated in three periods: before stimulation (called *pre*), immediately after the stimulus (called *post*) and the final period of recovery (called *final*). It is worth mentioning that the spontaneous neuronal activity recovers gradually and not suddenly, but these three intervals will be considered as a first approach. In the current experimental preparation, the spontaneous activity (measured as firing rate) of the studied neurons is fairly low, as can be seen in Figure 1.

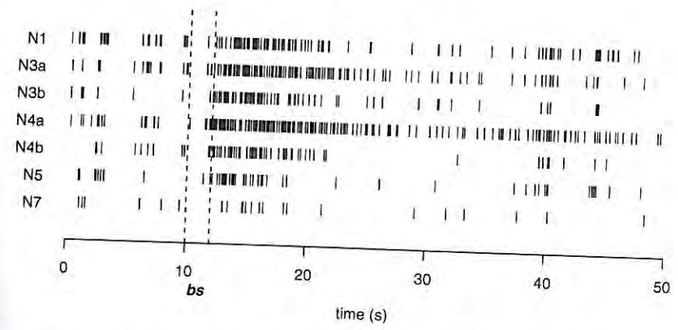

Fig. 1. Raster plots of 50 s recordings of one trial for each neuron. The short vertical lines represent the moments when the spikes occurred. The dotted lines represent the moments when the *bs* stimulus begins and ends.

3 Correlation Measures in Spike Trains

From a statistical viewpoint spike trains can be described as the sequence of AP generated by a point process, $\{T_i\}$, where T_i is the instant when the i-th AP occurred. In particular, if we have a T seconds duration neural activity recording and $(n+1)$ AP have occurred, we have a spike train $\{T_i\}_{i=1}^{n+1}$ with $0 \leq T_1 \leq T_2 \leq \ldots \leq T_{n+1} \leq T$. We can also define the inter spike intervals (ISI), $\{S_i\}_{i=1}^{n}$, as the elapsed times between consecutive spikes, where $S_i = T_{i+1} - T_i$, $i = 1, 2, \ldots, n$.

One useful feature to characterize neuronal activity is the correlation in spike trains. In the next subsections, two different methods to measure autocorrelation in spike trains are introduced.

3.1 Autocorrelation

Given the ISI of an observed spike train, $S_1, \ldots, S_n$, we can estimate the serial autocovariance function as

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-h} (S_{i+h} - \bar{S})(S_i - \bar{S}), \qquad 0 \leq h < n$$

where $\bar{S} = \frac{1}{n} \sum_{i=1}^{n} S_i$ is the sample mean. Then, the serial autocorrelation function is estimated by

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \qquad 0 \leq h < n\,.$$

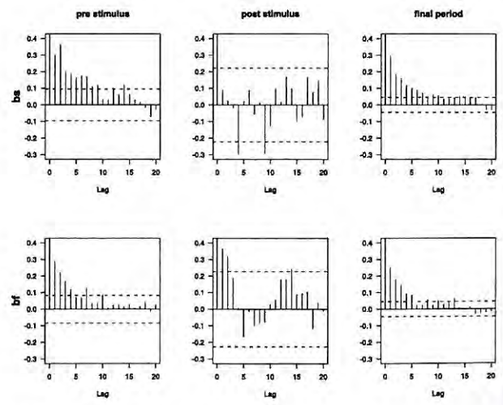

Fig. 2. Serial autocorrelograms for two trials of neuron N1 activity in each period. The dashed horizontal lines are the significance limits

In Figure 2 the autocorrelograms for two trials of neuron N1 can be observed. In this example, the structure of the autocorrelograms of the two trials are dissimilar. In the first one, there is serial autocorrelation of order up to nine for the *pre* period and there is no autocorrelation in the *post* period while in the second trial there is only correlation up to lower orders in *pre* and *final* and there exists a moderate autocorrelation in the *post* stage.

In general, autocorrelograms vary form trial to trial and there is no fixed structure for the autocorrelation of each neuron. Even though, it can be seen that the serial autocorrelation is low (most of the times not significant) in the *post* part of the recordings and that there exists autocorrelation of high orders in *pre* and *final*. A particular case is the one of neuron N5, which has a very low correlation of order one during spontaneous activity and absolutely no correlation during the effect of the stimulus. On the other hand, in the autocorrelograms for neuron N4a a very well defined alternation pattern between negative and positive coefficients can be observed. The autocorrelograms for neurons N4a and N5 can be seen in Figure 3.

For almost every neuron, the serial autocorrelation drops in the *post* period. This autocorrelation decrease is presumably due to the disruption of the spontaneous slow oscillatory activity induced by the electrical stimulation.

In [4] different pattern types of firing rates are discussed, such as, among others, cyclic rates or periods of local growth (or decrease). This patterns lead to different autocorrelogram shapes. Some of these patterns are shown by simulation studies and, for others, references are presented. One particular

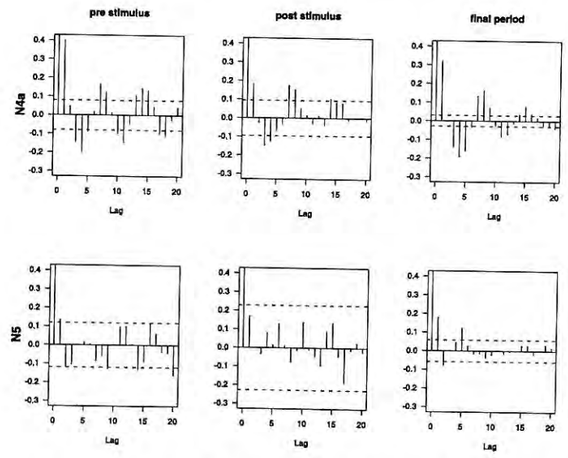

Fig. 3. Serial autocorrelograms for one trial of the *bs* stimulus of neurons N4a and N5 activity in each period. The dashed horizontal lines are the significance limits.

example of what is discussed in [4] is the pattern exhibited by neuron N4a: the alternation between negative and positive autocorrelation coefficients may proceed from an alternation between small and large ISIs.

3.2 Higher Order Interspike Autocorrelation

One may take into account, not only the elapsed times between consecutive firings but also the elapsed times between any two spikes. In fact this is the correlation measure mostly used in neuroscience.

The idea consists in breaking the total time of recording, T, in $Q = \left[\frac{T}{q}\right]+1$ intervals of length q. Let A_i be the i-th interval, $A_i = [(i-1)q, iq)$. Here it is convenient to note that, if only one spike is intended to fall in each interval, q must be sufficiently small, 1 ms for example. This is caused, for instance, by the refractory period that a neuron needs to recover before firing again. Let us define the new series $\{V_i\}_{i=1}^{Q}$:

$$V_i = \sum_{j=1}^{n+1} I(T_j \in A_i) \tag{1}$$

where $I(A)$ is the indicator function of the set A. We can estimate its autocovariance function by

$$\hat{\gamma}_V(h) = \frac{1}{Q}\sum_{i=1}^{Q-h}(V_{i+h}-\bar{V})(V_i-\bar{V}), \quad \bar{V} = \frac{1}{Q}\sum_{i=1}^{Q} V_i \,.$$

Now, if $h << Q$ the following approximations, $Q-h \approx Q$ and $\sum_{i=1}^{Q-h} V_{i+h}$ $\approx \sum_{i=1}^{Q} V_i \approx \sum_{i=1}^{Q-h} V_i$ can be used to obtain, after some algebra,

$$\hat{\gamma}_V(h) \approx \frac{1}{Q}\sum_{i=1}^{Q-h} V_{i+h}V_i - \bar{V}^2 \,. \tag{2}$$

Since $\bar{V}$ does not depend on h, we can concentrate in $\hat{\gamma}_V^*(h) = \hat{\gamma}_V(h) + \bar{V}^2$ $= \frac{1}{Q}\sum_{i=1}^{Q-h} V_{i+h}V_i$ without modifying the shape of the function $\hat{\gamma}_V(h)$.

Now,

$$V_{i+h}V_i = \sum_{j=1}^{n+1}\sum_{l=1}^{n+1} I(T_j \in A_i, T_l \in A_{i+h})$$

and then, $\hat{\gamma}_V^*(h)$, as a function of h, is the histogram of these frequencies, though divided by Q.

Observe that in the case of a small enough q (q=1 ms for example), $V_{i+h}V_i =$ 0, except when $V_{i+h} = V_i = 1$, in which case the product is 1. This is why, for each h:

$$\hat{\gamma}_V^*(h) = \frac{1}{Q}\sum_{i=1}^{Q-h} I((V_{i+h}, V_i) : (V_{i+h}, V_i) = (1,1)) \,,$$

which is easy to think in terms of time: $V_{i+h} = V_i = 1$ means that there are two spikes that are separated by a distance of, at least $h-1$ and at most $h+1$. From this follows that $Q\hat{\gamma}_V^*(h)$ counts the number of spike pairs (although some might be missing) that are separated by a distance between $h-1$ and $h+1$. This is the main idea to define the autocorrelation as it follows.

Higher order interspike autocorrelation, as it is used in neuroscience, is very similar to $\hat{\gamma}_V^*(h)$ but it is built in an alternative way. Actually, it is defined as the histogram of relative frequencies (or absolute sometimes) of the elapsed time between any two spikes of a train that do not surpass a certain $w_{\max}$ chosen by the researcher. This $w_{\max}$ is usually much smaller than T, which allows us to compare with the serial covariance function of $\{V_i\}_{i=1}^{Q}$, since the approximations in (2) are valid.

Given a spike train $\{T_i\}_{i=1}^{n+1}$, let the set of distances between any two spikes be $\{D_m\}_{m=1}^{M} = \{T_i - T_j / i,j \in \{1,\ldots,n+1\}, i \neq j\}$, such that $-w_{\max} \leq D_m \leq w_{\max}$. Moreover, we need to choose b, where $2b$ is the width of the histograms intervals. In this context, we define the higher order interspike autocorrelation (HOISA) of a spike train at the distance d, by:

$$\hat{g}(d) = \frac{1}{M}\sum_{m=1}^{M} I(d-b \leq D_m \leq d+b) \,.$$

Here b plays a similar role to q in (2) and, in fact, this histogram is very similar to that obtained from the serial autocovariance of the series $\{V_i\}_{i=1}^{Q}$. Some differences might arise from discretization and normalization. Interestingly, for $\hat{\gamma}_V^*(h)$ the discretization is done before calculating the distances, while in the definition of $\hat{g}(d)$ the discretization is carried out when constructing the histogram. On the other hand, to obtain $\hat{\gamma}_V^*(h)$ the absolute frequencies are divided by Q while for $\hat{g}(d)$ the denominator is M. Then the results should be almost proportional. In Figure 4 we can observe the degree of similarity of these histograms for three lengths of q and b, $q = b = 0.01, 0.1$ and 1 s for the activity in the *pre* period of one trial of neuron N1. The functions $\hat{\gamma}_V^*(h)$ and $\hat{g}(d)$ have been multiplied by Q and M respectively, so that the similarities are better shown up. The histograms are practically the same though there are some differences for the three sizes. These differences grow with the size of q and b and when $q = b = 1$ these are noticeable.

Note that, in fact, the HOISA is just an estimate of the probability density of time between any two spikes. To get a smoother estimate, a nonparametric kernel estimator is proposed:

$$\tilde{g}(d) = \frac{1}{Mh}\sum_{m=1}^{M} K\left(\frac{d-D_m}{h}\right) = \frac{1}{M}\sum_{m=1}^{M} K_h(d-D_m) \,,$$

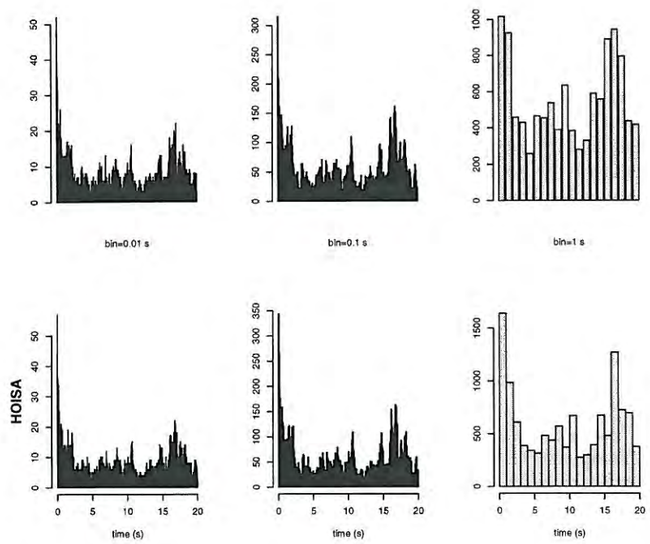

Fig. 4. Comparison between $Q\gamma_V^*(h)$ (top panel) and $M\hat{g}(d)$ (bottom panel) for three bin sizes, $q = b = 0.01, 0.1, 1$

where K is a kernel function and $h > 0$ a smoothing parameter (see [3], [5], [7] and [8] for details). We have used the gaussian kernel function and the Sheather and Jones "plug in" method for bandwidth selection (see [6]).

Figure 5 shows the autocorrelation for the *pre* and *post* periods for two trials, one for each stimulus, of neuron N1 activity. We have used a $w_{\max}$ of 10 s. In the *pre* period Figure 5 shows that, given that there is a spike in time t, there is a high probability density for another spike to occur within the next few seconds (less than 2 s). Also, differences between *pre* and *post* can be observed and also between the different trials in the *post* period.

Next, Figure 6 shows the estimates for a trial of each stimulus and for the three stages of the record of four different neurons. As we had observed before for neuron N1, in *pre* there is a high probability of two spikes occurring very close in time and there also exist some other probability peaks.

Many of the plots in Figure 6 exhibit secondary peaks. It is interesting to analyze what these secondary peaks mean. As indicated above, under sleep states or, as in this case, anesthesia, most cortical neurons display an oscillatory activity. Some rhythms have been characterized neurophysiologically in cats, as the slow rhythm (< 1 Hz), the delta rhythm (1–5 Hz) and the spindle oscillation (7–14 Hz). These rhythms are designated as *slow sleep*

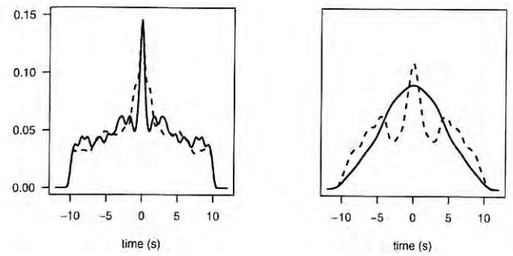

Fig. 5. Kernel estimates of the higher order interspike autocorrelation for one trial of stimulus *bs* (solid line) and one trial of stimulus *bf* (dashed line) of neuron N1 in the *pre* (left panel) and *post* (right panel) periods

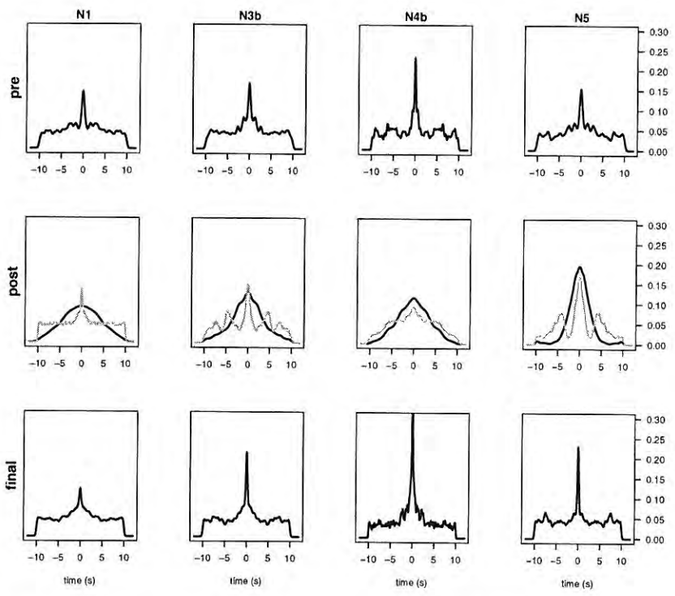

Fig. 6. Kernel estimates of the higher order interspike autocorrelation for one trial of each neuron in each period. Black lines correspond to trials of the *bs* stimulus and grey lines to the *bf* stimulus

oscillations. On the other hand, neuronal spike responses are grouped into what are called *bursts*. These features are sequences of action potentials fulfilling certain characteristics, including: a) consecutive spikes within a burst are not separated one from another in more than certain distance, and b) between one burst and another there is, at least, a certain distance. The distances in this definition may change for different areas of the cortex. If there is an oscillation, for example a delta oscillation of 2 Hz, what happens is that after a neuron generates a burst, it is quite likely that the next burst will occur after about 500 ms. In anesthetized cats it is common to record oscillations of about 0.1 Hz (belonging to the so termed *slow rhythm*). Thus, a slow oscillatory activity of 0.1 Hz could be the cause of the peaks at 10 s. If larger values of $w_{\max}$ were chosen, peaks at around 20 and 30 s could be observed in the HOISA.

Regarding the HOISA functions in the *post* stage. The differences between the autocorrelation functions are mainly found in their dispersion. For this period of the recorded trials, most of the histograms are unimodal, but there are some trials in which conspicuous secondary peaks can be observed; these are supposed to reflect stimulation-induced oscillations. In the *post* period it makes sense to compare the estimates obtained for each of the two stimuli. In several neurons, autocorrelations for the stimulus *bs* present more dispersion than those for the stimulus *bf*.

The estimates of the autocorrelation function for the *final* stage of the study are very similar to the corresponding ones of the *pre* condition. The main peak of the probability density remains at zero. There are also other peaks as in *pre*. In the *post* period, distances between spikes were mainly small but when the effect of the stimulus is over, the distances return to the behavior they had before the stimulus was applied.

These correlation measures, counting the distances along the entire train, do not take into account the possible lack of stationarity on trains and therefore assume that the correlation is stationary. Often, this stationarity is not easy to justify. When neurons are under the effect of a stimulus, they can adapt to it over time, or they can lose stationarity because the stimulus vary in time and the neuronal response varies with it. To check the lack of stationarity and see how the autocorrelation changes over time, the HOISA functions can be computed using sliding windows. Thus, autocorrelation is calculated at each instant using information of a neighborhood of each point. The problem with this approach is that it takes a lot of data for the estimates to be accurate. From the neurobiological perspective it is interesting to calculate these functions in the *post* period, to reveal the autocorrelation dynamics after the stimulus is applied. However, with the amount of data we had for this analysis no accurate estimates could be computed. Instead, in Figure 7 we see an example of how this functions behave in the *pre* stage, where they could, or not, be stationary.

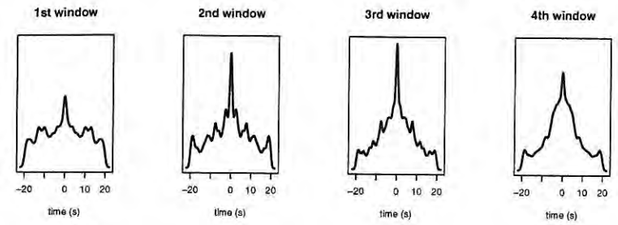

Fig. 7. HOISA functions in the *pre* part of the first trial of neuron N1. A sliding window of 48 s has been used centered at times: 24 s, 48 s, 72 s and 96 s.

4 Testing Independence for Interspike Intervals

In this section we will study the existence of dependence among the elapsed times between consecutive spikes. So, we will test the null hypothesis H_0 : the ISIs are independent, versus the alternative H_1 : they are not independent.

Two different tests will be proposed. If the ISIs are dependent, this situation will influence the shape of the HOISA. These functions will be used to build the first test. The estimated autocorrelation function for the original train will be compared with another one obtained from independent spike trains. On the other hand, the Kolmogorov-Smirnov goodness-of-fit test will be used to compare the distribution of the elapsed times between consecutive spikes in the original train with the distribution of the times of independent trains.

To obtain the sample of independent ISIs, a random shuffle is performed in the original ISIs. A new sample $\{S_i^*\}_{i=1}^n$ is obtained from $\{S_i\}_{i=1}^n$, destroying all the possible serial dependence but preserving any other possible features. With this new sample a new spike train is built: $T_1^* = 0$ y $T_i^* = \sum_{j=1}^n S_j^*$, whose times between consecutive spikes are independent. The differences between the HOISA function of this independent train and the one of the original train will show how far from independence the train under study is. In [4] this methodologies for independence tests are discussed.

The first hypothesis test is carried out as follows. The HOISA function of a registered spike train will be compared with the one obtained from a shuffled train. More specifically, N shuffled trains are used, their HOISA functions are computed and averaged to avoid falling in a case that is not representative. This *average HOISA function* is denoted $\bar{g}(t)$. The test statistic is defined as the L_1 distance:

$$T_{HOISA} = \int |\tilde{g}(x) - \bar{g}(x)| \mathrm{d}x \ .$$

H_0 will be rejected for large values of T_{HOISA}.

For the second test, the empirical distributions of the D_m for both populations, the original, $\tilde{F}$, and the shuffled one, which we call $\bar{F}$, are needed. To

estimate the second one, as before, several shuffled trains are used, say N. From each shuffled train, the set of distances between spikes is constructed, and $\bar{F}$ is defined as the empirical distribution of the sample of all these N sets put together. Then, the Kolmogorov-Smirnov test statistic is used:

$$T_{KS} = \sup_x |\tilde{F}(x) - \bar{F}(x)| \ .$$

To calibrate the distributions of the test statistics a bootstrap method is proposed. The steps for the first test are the following:

1. Sample from $\{s_i = t_{i+1} - t_i\}_{i=1}^n$ to obtain a resample $\{s_i^*\}_{i=1}^n$ of distances between consecutive spikes and build a bootstrap train: $t_1^* = 0$, and $t_i^* = \sum_{j=1}^{i-1} s_j^*$, for $i = 2, \ldots, n+1$.
2. Calculate $\tilde{g}^*(t)$ for this bootstrap train.
3. Resample N times from s^* to obtain: $s^{**(i)}$, $i = 1, \ldots, N$ as before. Build $t^{**(i)}$ as in Step 1 and calculate $\tilde{g}^{**(i)}$ for each train $t^{**(i)}$. Then define $\bar{g}^* = \frac{1}{N} \sum_{i=1}^N \tilde{g}^{**(i)}$.
4. Obtain $T^*_{HOISA} = \int |\tilde{g}^* - \bar{g}^*|$.
5. Repeat Steps 1-4 B times to get $T^*_{HOISA,1}, \ldots, T^*_{HOISA,B}$ and use them to estimate the desired quantiles of the T_{HOISA} distribution or the p-value for T^{obs}_{HOISA}.

For the Kolmogorov-Smirnov test the procedure is very similar:

1. Build the independent spike train t^* as before.
2. Calculate the distances between consecutive spikes for the bootstrap train: $\{d_m^*\}_{m=1}^M$.
3. Resample N times from t^*, to build N trains $\{t^{**(j)}, j = 1, \ldots, N\}$ and for each train build the set of distances: $\{d_m^{**(j)}\}_{m=1}^{M_j}$.
4. Calculate T^*_{KS} as the Kolmogorov-Smirnov statistic for the samples $\{d_m^*\}_{m=1}^M$ y $(d_1^{**(1)}, \ldots, d_{M_1}^{**(1)}, d_1^{**(2)}, \ldots, d_{M_2}^{**(2)} \ldots, d_1^{**(N)}, \ldots, d_{M_N}^{**(N)})$.
5. Repeat Steps 1-4 B times to obtain $T^*_{KS,1}, \ldots, T^*_{KS,B}$ and use them to estimate the desired quantiles of the T_{KS} distribution or the p-value for T^{obs}_{KS}.

In general, differences can be observed between the HOISA function of the original train and the one obtained with the resamples. Roughly speaking, the density of the resampled data is more uniformly distributed and then the main peak is lower than in the case of the real data. It is also very common the absence of secondary peaks in the HOISA functions of the resampled trains.

In Table 1 the results of the tests for four neurons, N1, N3a, N3b and N4b and three different recordings (one in the *pre* part and two in the *post* part, one for each stimulus) can be observed. The p-values obtained with each test were calculated using the bootstrap method described above. A total number of 500 bootstrap resamples and $N = 80$ shuffles were used for each bootstrap train in the *pre* part and $N = 100$ in the *post* part. Also, a Ljung-Box (T_{LB}) test was implemented to compare the results.

Table 1. p-values for the independence tests T_{HOISA}, T_{KS} and T_{LB}, constructed using the distances between two spikes

neuron	N1			N3a			N3b			N4b		
period	*pre*	*post*		*pre*	*post*		*pre*	*post*		*pre*	*post*	
stimulus		*bs*	*bf*		*bs*	*bf*		*bs*	*bf*		*bs*	*bf*
T_{HOISA}	0.000	0.080	0.204	0.006	0.004	0.126	0.000	0.006	0.308	0.000	0.254	0.574
T_{KS}	0.002	0.036	0.668	0.266	0.002	0.456	0.002	0.004	0.634	0.001	0.148	0.514
T_{LB}	0.000	0.205	0.012	0.000	0.000	0.018	0.000	0.000	0.980	0.002	0.320	0.911

These results show that, in the *pre* period, the distances between consecutive spikes are not independent (but one case: N3a KS test). This fact is not true in most of the cases for the *post* period. The null hypothesis is rejected in the *post* part for the *bs* stimulus in neurons N3a and N3b (and for *bs* in N1 using the KS test and in *bf* in N1 and N3a using the Ljung-Box test) but it is not rejected for the *bf* stimulus, showing a difference between stimuli.

Figure 8 shows the HOISA functions of three original trains, both *pre* and *post* (*bs* stimulus) periods, and the average HOISA function for the independent case, averaged over 100 shuffles of the original train. In Figure 8 it is easy to recognize the cases where independence is rejected (neurons N1, N3b and the *pre* period of neuron N4b) and the one in which it is not (*post* of neuron N4b).

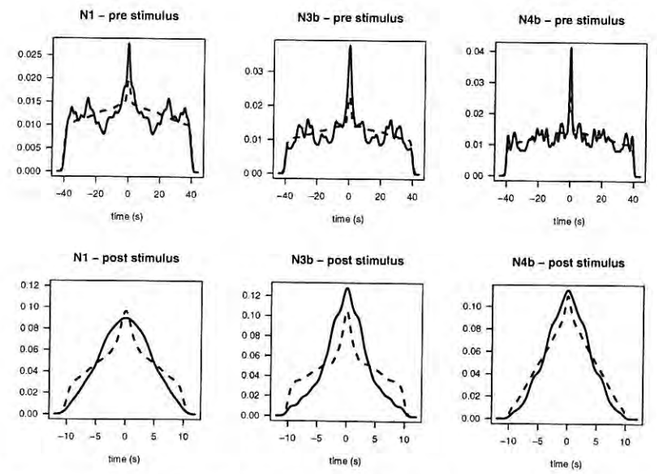

Fig. 8. Comparison of the HOISA function for the original trains (solid line) and the average HOISA function for independent trains (dashed line). First trial of neurons N1, N3b and N4b (*bs* stimulus).

5 Conclusions

Two correlation measures for spike trains have been introduced. First of all, the serial autocovariance has been discussed. On the other hand, the higher order interspike autocorrelation (HOISA) was presented. This autocorrelation measure is commonly used in neuroscience. The spontaneous activity of neurons is characterized by the existence of dependence among spikes. In this work, a test for independence based on the HOISA function was proposed. As this function is constructed in the basis of a histogram, another statistic, based on the empiric distribution function, was discussed. The distribution of these statistics under the null hypothesis was calibrated with a bootstrap procedure. Finally, a Ljung-Box statistic was also used for comparison. This last statistic has the inconvenience of being based on the serial autocorrelation which, varies very much from one trial to another. In general, it can be observed that dependence exists in the *pre* part, reflecting the highly synchronized neuronal oscillatory activity. This dependence is present for some neurons after the *bs* stimulus while it does not appear after the *bf* stimulus for most of the neurons. In some cases, the T_{KS} and T_{LB} statistics present values that are not consistent with the ones obtained with the other tests. This does not happen with the T_{HOISA} statistic, what makes it more robust. Our results indicate that the HOISA-based test for independence is a useful method for the characterization and analysis of the dynamics of the neuronal oscillatory activity.

References

1. Brown, N.E., Kass, R.E., Mitra, P.: Multiple neural spike train data analysis: state-of-the-art and future challenges. Nature Neurosci. 7(5), 456–461 (2004)
2. Dayan, P., Abbott, L.F.: Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. The MIT Press, Cambridge (2001)
3. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076 (1962)
4. Perkel, D.H., Gerstein, G.L., Moore, G.P.: Neuronal spike trains and stochastic processes. I. The single spike train. Biophys. J. 7(4), 391–418 (1967)
5. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832–837 (1956)
6. Sheather, S.J., Jones, M.C.: A reliable data-based bandwidth selection method for kernel density-estimation. J. Royal Statist. Soc. Ser. B 53(3), 683–690 (1991)
7. Silverman, B.W.: Density Estimation. Chapman & Hall, London (1986)
8. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman & Hall, London (1995)

Multiple Comparison of Change Trends in Cancer Mortality/Incidence Rates Taking with Overlapping Regions and Time-Periods*

Nirian Martín[1] and Yi Li[2]

[1] Dept. Statistics, Carlos III University of Madrid, Spain
nirian.martin@uc3m.es
[2] Dept. Biostatistics – Harvard School of Public Health and
Dep. Biostatistics & Computational Biology – Dana Farber Cancer Institute, USA
yili@jimmy.harvard.edu

Summary. When analyzing trends in cancer rates, it is common to rely on the so-called Annual Percent Change (APC). For dealing with such a measure of trend, directly age-adjusted rates are usually considered. Classical methods such as pooled t-tests are often applied for comparing APCs of two groups of individuals in a simple way under independence assumption. In practice, it is quite common to find groups of interest for which the independence assumption fails because their regions or periods of time overlap. No one of the papers that deal this problem consider the case where more than two APCs are compared. In this work we propose a Wald-type test-statistic which is not difficult to compute once we provide the estimators of two or more APCs to be compared. These estimators are the minimum power divergence estimators that cover as special case those obtained by maximum likelihood.

1 Introduction

According to the World Health Statistics 2009, published by WHO, the different types of cancer jointly cardiovascular diseases constituted the main causes of death during year 2004, specially in upper-middle and high income countries. It is well known that as income level increases, the percentage of non-communicable deaths (NCDs) increases, in other words, as people live longer, the risk of NCD grows. In such a study it is shown that in high-income countries during year 2004, 77% of all the deaths were NCDs. Cardiovascular diseases and cancers, as main part of NCDs, provided values of

* This work is related to the stay of Nirian Martín in Harvard University (September 2008 - August 2009), supported by the Real Colegio Complutense. She met Marisa just the first day after arriving in Spain, September 1. In the last conversation Marisa told Nirian that both will meet soon again. Due to her optimistic character it was impossible to imagine they both could not meet again. This paper has been written in memory to Marisa.