BIOMETRIE UND MEDIZINISCHE INFORMATIK
GREIFSWALDER SEMINARBERICHTE

Heft 15

Christos P. Kitsos
Alexandros G. Rigas
Karl-Ernst Biebler (Ed.)

Cancer Risk Assessment

Selected papers from ICCRA3

SHAKER
VERLAG

A non-parametric Test Based on Cumulative Periodograms for Comparing Two Neural Spike Trains

Alexandros G. Rigas
Democritus University
of Thrace, Greece
rigas@ee.duth.gr

Vassilios G. Vassiliadis
Democritus University
of Thrace, Greece
bvasil@ee.duth.gr

Abstract. In two recent papers by Rigas and Vassiliadis (2007) and Vassiliadis and Rigas (2009) semi-parametric tests were proposed for comparing the estimated spectra of two stationary point processes. Neural spike trains are usually described as realizations of stationary point processes. The tests are useful in assessing if the information contained in the one neural spike train is the same or differs from the other and in which band of frequencies. In this work, we present a non-parametric approach using the cumulative periodograms and a graphical method for testing the hypothesis that the underlying spectra are equal. An illustrative example from the field of Neurophysiology is presented where the neuromuscular system of the muscle spindle is affected simultaneously by the presence of a gamma and an alpha motoneuron. It is shown that the response of the muscle spindle can be separated in two parts (0-10) Hz and (10-100) Hz. The first part is related to the effect of the gamma motoneuron, while the second part to the effect of the alpha motoneuron.

1. Introduction

Neural spike trains are usually considered as realizations of a point process. We denote by $\{N_k(t)\}$, $k=1,2$ two stationary point processes (SPPs) observed in the interval $(0,T]$, where $N_k(t)$ is the (cumulative) spike count of the type-k events in the interval $(0,T]$. The point processes are assumed to be orderly and obey a strong mixing condition (Brillinger (1975), Daley and Vere-Jones (1988)). Let p_k and $p_{kk}(u)$ denote the mean intensity and the second-order cumulant density of $\{N_k(t)\}$ respectively ($k=1,2$). Then, the second-order density function (SDF) of $\{N_k(t)\}$ is defined by

$$f_{kk}(\lambda)=(2\pi)^{-1}p_k+(2\pi)^{-1}\int_{-\infty}^{\infty}q_{kk}(u)\exp(-i\lambda u),\ -\infty<\lambda<\infty \quad (1.1)$$

For large λ, $f_{kk}(\lambda)$ tends to a constant value, that is

$$\lim_{\lambda\to\infty} f_{kk}(\lambda)=\frac{p_k}{2\pi},\ (k=1,2) \quad (1.2)$$

This constant value corresponds to the SDF of a Poisson point process with mean intensity p_k. The SDF exists when $\int_{-\infty}^{\infty}|q_{kk}(u)|\,du<\infty$. More details about the SPPs and the parameters defined above can be found in Bartlett (1963) and Cox and Isham (1980).

The test we intend to discuss is based on the SDFs and can be specified as follows:

$$\begin{matrix} H_0: f_{11}(\lambda)=cf_{22}(\lambda) \\ H_1: f_{11}(\lambda)\neq cf_{22}(\lambda) \end{matrix},\ \lambda_L<\lambda<\lambda_H, \quad (1.3)$$

where c is a constant and H_1 implies that we wish to detect only shape differences between the underlying spectra. The values of λ are restricted in the interval (λ_L,λ_H) since we are interested in examining the information, which is contained in a band of frequencies in the SDF.

In Coates and Diggle (1986) various periodogram-based tests are presented for the hypothesis that two independent time series are realizations of the same stationary process. In these tests, the null hypothesis was tested against the particular alternative:

$$f_{11}(\lambda)=\exp(a+b\lambda+c\lambda^2)\cdot f_{22}(\lambda) \quad (1.4)$$

The test by Coates and Diggle, which is based on the likelihood ratio test, has been extended in Rigas and Vassiliadis (2007) by smoothing the periodogram statistic using two different approaches. It was shown that the new test was more powerful and in our examples the relation of the examined power spectra is of the quadratic form given by (1.4) which implies that a shape difference exists. An alternative approach has been presented in which the (smoothed) periodogram-based estimates are modeled as generalized linear models (GLMs) with errors gamma-distributed variates. In a recent paper Vassiliadis and Rigas (2009) proposed an alternative test which is based on the estimate of the SDF obtained by using convergence factors and estimates of the time parameters p_N and $q_{NN}(u)$.Then the quasi-likelihood atio test is used in order to test the hypothesis (1.4). In Diggle and Fisher (1991) a non-parametric approach is described and a simple graphical method for comparing two periodograms is also presented.

In this work we present an application of the non-parametric test described in Diggle and Fisher (1991) by using the modified periodogram statistics of SPPs. This is illustrated by using an example from the field of neurophysiology where the neuromuscular system of the muscle spindle is affected simultaneously by the presence of a gamma and an alpha motoneuron.

2. Smoothing procedure

Let $\{N_k(t_n);\ t_n = nb;\ n = 1,\ldots,T/b\}$ be two ($k = 1,2$) discretized SPPs which are orderly and satisfy a strong mixing condition. The discretized SPPs are sampled from continuous ones with the same sampling rate $1/b$. The second-order modified periodogram statistic of $\{N_k(t)\}$ is defined by

$$\hat{I}_{kk}^{(T)}(\lambda) = (2\pi T)^{-1} \left|\hat{d}_k^{(T)}(\lambda)\right|^2,\ \lambda \neq 0(\text{mod}\, 2\pi), \tag{2.1}$$

where $\hat{d}_k^{(T)}(\lambda)$ is the finite mean corrected Fourier-Stieltjes transform of the increments of the k-th point process, given by

$$\hat{d}_k^{(T)}(\lambda) = \int_0^T \exp(-i\lambda t)\left[dN(t) - \hat{p}_k dt\right] \tag{2.2}$$

In practice (2.2) can be approximated by the following expression

$$\hat{d}_k^{(T)}(\lambda) \approx \sum_{j=0}^{T/b-1} \exp(-i\lambda t_j)\left[N(t_j + b) - N(t_j) - \hat{p}_k b\right], \tag{2.3}$$

where $\hat{p}_k = N_k(T)/T$ is an estimate of the mean intensity. By $N_k(T)$ we denote the number of events of $\{N_k(t)\}$ in the interval $(0,T]$. For more details about this approach refer to Rigas (1992). This discretization approach permits the use of the FFT algorithm in the computation of the modified periodogram statistic. It can be proved that

$$\hat{I}_{kk}^{(T)}(\lambda) \sim f_{kk}(\lambda) X_2^2 / 2 \tag{2.4}$$

Moreover for $j \neq i$, $\hat{I}_{kk}^{(T)}(\lambda_i)$ and $\hat{I}_{kk}^{(T)}(\lambda_j)$ are asymptotically independent. These results are discussed in Brillinger (1978) and Diggle (1990).

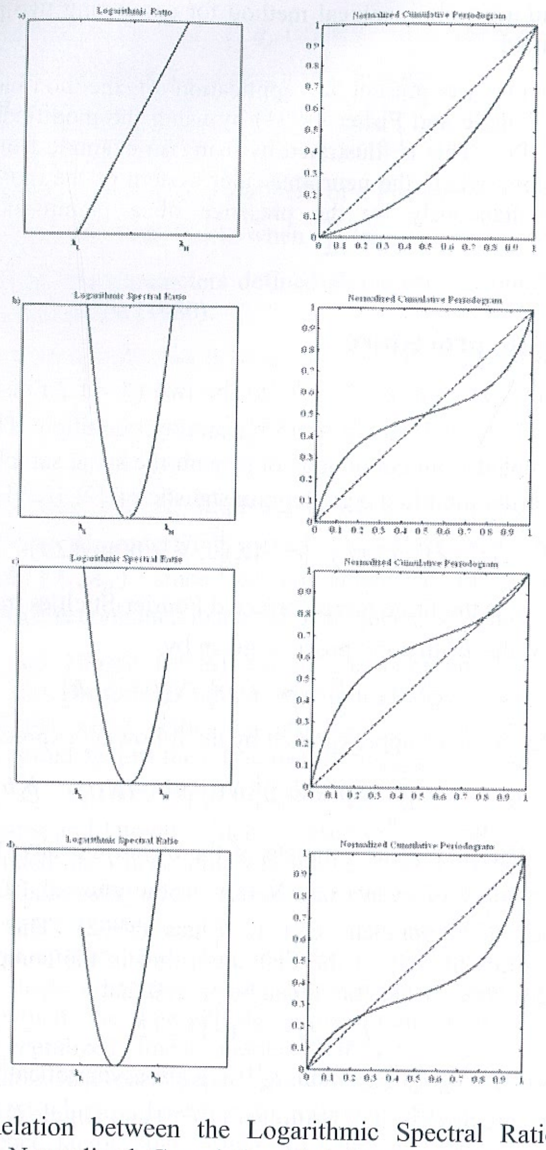

Figure 2.1 Relation between the Logarithmic Spectral Ratio and the corresponding Normalized Cumulative Periodogram scatter plot in four cases: a) linear Logarithmic Spectral Ratio, b) c) and d) quadratic Logarithmic Spectral Ratio.

For each of the two discretized SPPs ($k=1,2$) the corresponding estimated normalized cumulative periodogram (NCP) is given by

$$\hat{F}_{kk}(\lambda_j)=\sum_{i=1}^{j}\hat{I}_{kk}^{(T)}(\lambda_i)\Big/\sum_{i=1}^{n}\hat{I}_{kk}^{(T)}(\lambda_i) \tag{2.5}$$

This definition of the normalized cumulative periodogram is discussed in Diggle and Fisher (1991). In order to compare the spectra of the two SPPs the idea is to compare their estimated spectra via a plot of $\hat{F}_{11}(\lambda_j)$ against $\hat{F}_{22}(\lambda_j)$ which are obtained from (2.5). The use of the NCP implies that we wish to detect only shape differences between the two underlying spectra and thus it is appropriate for testing H_1 of (1.3). It is obvious that testing for shape differences is more general than revealing a quadratic model such the one given by (1.4). Nevertheless, the NCP plot behaves in a characteristic manner when certain linear or quadratic models are assumed. The NCP plot under the null hypothesis is expected to be approximately a straight line. Figure 2.1a demonstrates the behavior of the NCP plot under the hypothesis that the logarithm of the spectral ratio is given by a linear model. Figures 2.1b, 2.1c and 2.1d show the behavior of the NCP under the hypothesis that the logarithm of the spectral ratio is given by three different quadratic models.

We now describe a way of formulating a formal test in order to enforce the NCP plot and to decide whether possible deviations from the straight line are statistically significant. Let $\boldsymbol{F}_k=\left[F_{kk}(\lambda_1),\ldots,F_{kk}(\lambda_m)\right]$, $k=1,2$ be two vectors containing the theoretical NCPs of the two SPPs where a certain renumbering has been made implying that $\lambda_1,\ldots,\lambda_m$ are the m frequencies that lie in the desired frequency band (λ_L,λ_H). Then under H_0, $\boldsymbol{F}_1=\boldsymbol{F}_2$ and the deviations from it can be described using any measure d of the distance between $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$. It is obvious that the distribution of the random variable d, under H_0, is not known a priori but it is not affected by any of 2^m possible interchanges between $f_{11}(\lambda_j)$ and $f_{22}(\lambda_j)$, $j=1,\ldots,m$. In practice, we can approximate the distribution of d by calculating $d_2,d_3,\ldots,d_s$ for some large number $s-1$ of random interchanges of spectral ordinates at the each frequency. If by d_1 is denoted the observed distance, then the significance probability can be thought as the proportion of the values $d_2, d_3, \ldots, d_s$ that are at least as large as d_1 (Diggle and Fisher (1991)). For the rest of the paper, as a measure of distance d we choose the Kolmogorov-Smirnov distance, i.e.

$$D_m = \sup\left|F_{11}(\lambda_j) - F_{22}(\lambda_j)\right| \qquad (2.5)$$

to be consistent with previous works.

3. Neurophysiological Example

An example from the field of neurophysiology is now presented in which the above procedure is illustrated. Three data sets describing the response of the neurophysiological system called muscle spindle are used in the following cases:

i) The muscle spindle is influenced by a gamma motoneuron (gamma stimulation) (G)

ii) The muscle spindle is influenced by an alpha motoneuron (alpha stimulation) (A)

iii) The muscle spindle is under the simultaneous influence of a gamma and an alpha motoneuron (GA).

The responses of the muscle spindle are sequences of nerve impulses and are considered as realizations of SPPs, observed in a time interval of length $T = 11360ms$. Our interest here is to examine if the information carried through the Ia sensory axon from the muscle spindle to the spinal cord under the simultaneous influence of the two stimuli can be divided into different bands of frequencies which are related to the effects of each stimulus respectively. The condition of orderliness is satisfied since the sampling period is chosen to be $1ms$. The first band (0-10)Hz corresponds to the influence of the gamma motoneuron. The second band (10-100)Hz corresponds to the influence of the alpha motoneuron.

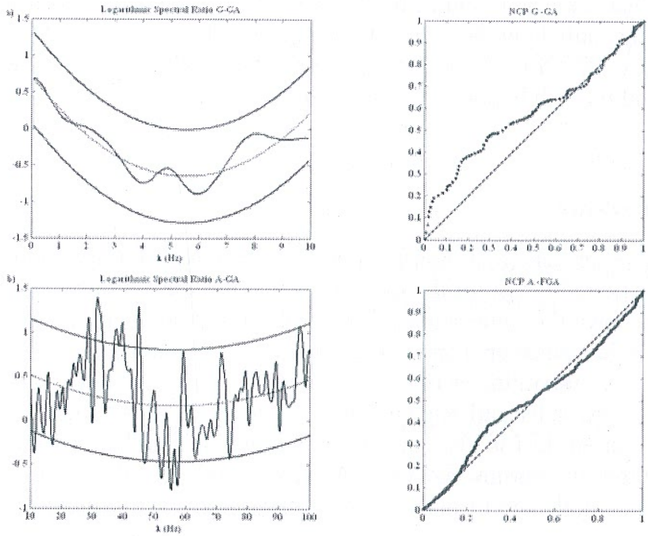

Figure 3.1 The Logarithmic Ratio of estimated spectra and the corresponding Normalized Cumulative Periodogram scatter plot a) of the (G) and (GA) cases in the first frequency band and b) of the (A) and (GA) cases in the second frequency band

In Figure 3.1a the logarithmic ratio of the estimated spectra of the (G) and (GA) cases is shown in the first frequency band. In Figure 3.1b the logarithmic ratio of the estimated spectra of the (A) and (GA) cases is shown in the second frequency band. The estimates in both the left plots are obtained by using a Tukey convergence factor following the method described in Vassiliadis and Rigas (2009). The quadratic form of the ratio suggests a significant difference in the shapes of the spectra. The dotted line corresponds to the estimate $\hat{a}+\hat{b}\lambda+\hat{c}\lambda^2$ for the hypothesis given (1.4). The solid lines correspond to the 95% approximate confidence limits. In Figures 3.1a and 3.1b (right plots) the normalized cumulative periodogram scatter plots are presented for each case. These also suggest that there is a significant difference in the shapes of the spectra. The shape of the NCP plot also suggests a quadratic or a more complicated form for the logarithmic ratio when it is compared with the plots presented in Figure 2.1. When the formal test was performed for the first case the Kolmogorov-

Smirnov distance was found to be 0.201 with a p-value equal to 0.045 and therefore the null hypothesis must be rejected. For the second case, the Kolmogorov-Smirnov distance was found to be 0.090 with a p-value equal to 0.003 and the null hypothesis must also be rejected.

4. Conclusions

We have presented a non-parametric test based on the normalized cumulative periodogram for the hypothesis that the spectra of two stationary spike trains have the same shape. This graphical method is useful because it helps to decide whether a quadratic model is appropriate to describe the logarithm spectral ratio, as the one presented be Rigas and Vassiliadis (2007) or Vassiliadis and Rigas (2009), or a more complicated model is needed. A formal test is also used by combining a simulation method and the Kolmogorov Smirnov distance. A power study for this test has been presented in Diggle and Fisher (1991) for certain time series models.

A neurophysiological example is discussed where the information carried to the from the muscle spindle to the spinal cord under the simultaneous influence of an alpha and a gamma motoneuron can be divided into two different frequency bands, one in the range (0-10)Hz and the other in the range (10-100)Hz. It is shown that the first range is connected with the influence of the gamma motoneuron and the second one with the influence of the alpha motoneuron. In both cases the spectra shapes are different and a more general relation exists between spectra. In addition cases the NCP plots suggest a quadratic or more complicated relation between the spectra.

Acknowledgments

The authors are most grateful to Professor G.P. Moore and Professor J.R. Rosenberg for providing the data sets.

References

[1] M.S. Bartlett. The spectral analysis of point processes, J. R. Statist. Soc. B25:264-296, 1963.

[2] D.R. Brillinger. Estimation of product densities. In Proceedings of the Computer Science and Statistics: 8th Annual Symposium on Interface, UCLA, 431-438, 1975.

[3] D.R. Brillinger. Comparative aspects of the study of ordinary time series and of point processes. In P.R. Krishnaiah (Ed.), Developments in Statistics, vol. 1. Academic Press, 33–134, 1978.

[4] D.S. Coates and P.J. Diggle. Tests for comparing two estimated spectral densities. J. Time Ser. Anal. 7(1):7-20, 1986.

[5] D.R. Cox and V. Isham. Point Processes. Chapman & Hall, 1980.

[6] D.J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes. Springer, 1988.

[7] P.J. Diggle. Time Series: A Biostatistical Introduction. Clarendon Press, 1990.

[8] P.J. Diggle and N.I. Fisher. Nonoparametric comparison of cumulative periodograms. Appl. Statist. 40(3):423-434, 1991.

[9] A.G. Rigas. Spectral analysis of stationary point processes using the fast Fourier transform algorithm. J. Time Ser. Anal. 13 (5):441-450, 1992.

[10] A.G. Rigas and V.G. Vassiliadis. A semi-parametric approach for comparing the estimated spectra of two stationary processes. Math. Biosci. 210:361-377, 2007.

[11] V.G. Vassiliadis and A.G. Rigas. A semi-parametric test based on quasi-likelihood approach for comparing the estimated spectra of two neural spike trains. Neurocomputing “doi:10.1016/j.neucom. 2009.02.008”, 2009.