STAR: Spike Train Analysis with R

Christophe Pouzat[1], Antoine Chaffiol[2] and Chong Gu[3]

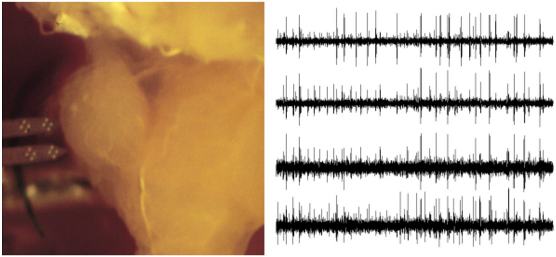
[1]Paris-Descartes University, Paris, France
christophe.pouzat@gmail.com
[2]INRA UMR 1272, Versailles, France
achaffiol@versailles.inra.fr
[3]Purdue University, West Lafayette, USA
chong@stat.purdue.edu

useR 2009: July 9

Outline

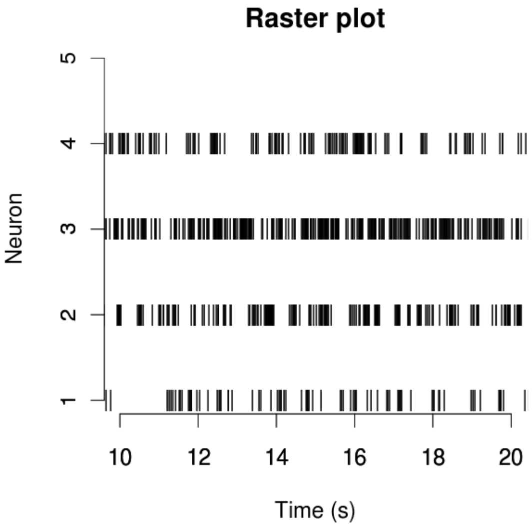
In vivo multi-electrodes recordings from insects

"From the outside" the neuronal activity appears as brief electrical impulses: the action potentials or spikes.



Left, the brain and the recording probe with 16 electrodes (bright spots). Width of one probe shank: 80 μm. Right, 1 sec of raw data from 4 electrodes. The local extrema are the action potentials.

Spike trains



After a rather heavy pre-processing stage called spike sorting spike trains are obtained.

Studying spike trains *per se*

- A central working hypothesis of systems neuroscience is that action potential or spike occurrence times, as opposed to spike waveforms, are the sole information carrier between brain regions.

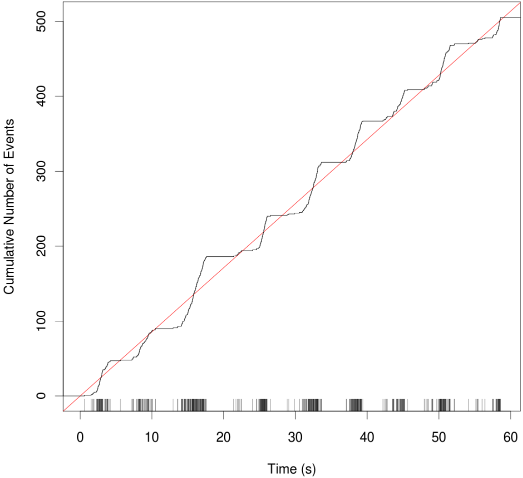
Studying spike trains *per se*

- A central working hypothesis of systems neuroscience is that action potential or spike occurrence times, as opposed to spike waveforms, are the sole information carrier between brain regions.
- This hypothesis legitimates and leads to the study of spike trains *per se*.

Studying spike trains *per se*

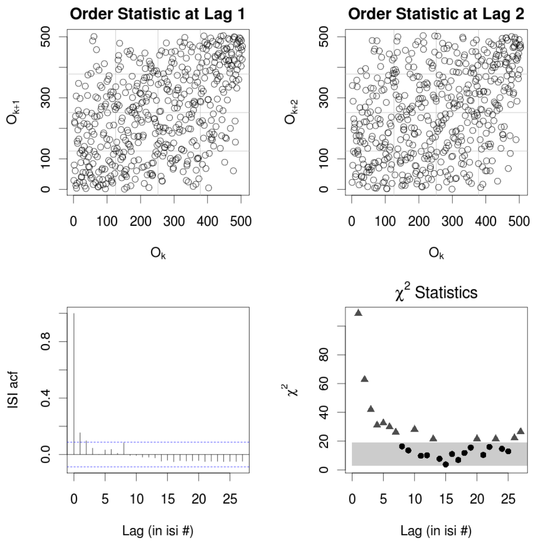
- A central working hypothesis of systems neuroscience is that action potential or spike occurrence times, as opposed to spike waveforms, are the sole information carrier between brain regions.
- This hypothesis legitimates and leads to the study of spike trains *per se*.
- It also encourages the development of models whose goal is to predict the probability of occurrence of a spike at a given time, without necessarily considering the biophysical spike generation mechanisms.

Spike trains are not Poisson processes



The "raw data" of one bursty neuron of the cockroach antennal lobe. 1 minute of spontaneous activity.

Spike trains are not Renewal processes



Some “renewal tests” applied to the previous data.

A counting process formalism (1)

Probabilists and Statisticians working on series of events whose only (or most prominent) feature is there occurrence time (car accidents, earthquakes) use a formalism based on the following three quantities (Brillinger, 1988, Biol Cybern 59:189).

- Counting Process: For points $\{t_j\}$ randomly scattered along a line, the counting process $N(t)$ gives the number of points observed in the interval $(0, t]$:

$$N(t) = \sharp\{t_j \text{ with } 0 < t_j \leq t\}$$

where $\sharp$ stands for the cardinality (number of elements) of a set.

A counting process formalism (2)

- History: The history, $\mathcal{H}_t$, consists of the variates determined up to and including time t that are necessary to describe the evolution of the counting process.
- Conditional Intensity: For the process N and history $\mathcal{H}_t$, the conditional intensity at time t is defined as:

$$\lambda(t \mid \mathcal{H}_t) = \lim_{h \downarrow 0} \frac{\text{Prob}\{\text{event } \in (t, t+h] \mid \mathcal{H}_t\}}{h}$$

for small h one has the interpretation:

$$\text{Prob}\{\text{event } \in (t, t+h] \mid \mathcal{H}_t\} \approx \lambda(t \mid \mathcal{H}_t)\, h$$

Goodness of fit tests for counting processes

- All goodness of fit tests derive from a mapping or a "time transformation" of the observed process realization.
- Namely one introduces the integrated conditional intensity :

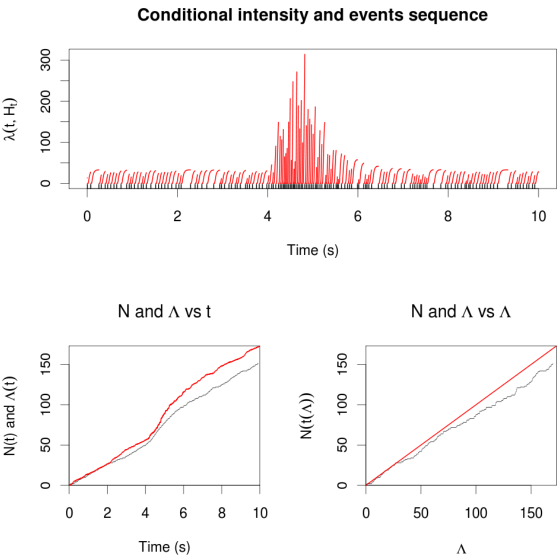
$$\Lambda(t) = \int_0^t \lambda(u \mid \mathcal{H}_u) \, du$$

- If Λ is correct it is not hard to show that the process defined by :

$$\{t_1, \ldots, t_n\} \mapsto \{\Lambda(t_1), \ldots, \Lambda(t_n)\}$$

  is a Poisson process with rate 1.

Time transformation illustrated



An illustration with simulated data.

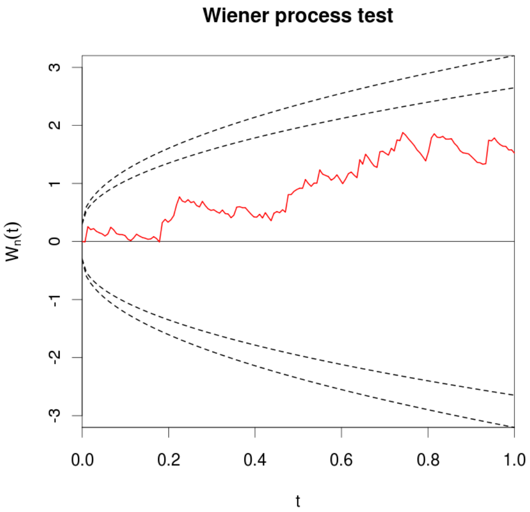
A goodness of fit test based on Donsker's theorem

- Y Ogata (1988, JASA, 83:9) introduced several procedures testing the time transformed event sequence against the uniform Poisson hypothesis.
- We propose an additional test built as follows :

$$\begin{aligned} X_j &= \Lambda(t_{j+1}) - \Lambda(t_j) - 1 \\ S_m &= \sum_{j=1}^{m} X_j \\ W_n(t) &= S_{\lfloor nt \rfloor}/\sqrt{n} \end{aligned}$$

- Donsker's theorem (Billingsley, 1999, pp 86-91) states that if Λ is correct then W_n converges weakly to a standard Wiener process.
- We therefore test if the observed W_n is within the tight confidence bands obtained by Kendall et al (2007, Statist Comput 17:1) for standard Wiener processes.

Illustration of the proposed test



The proposed test applied to the simulated data. The boundaries have the form: $f(x; a, b) = a + b\sqrt{x}$.

Where Are We?

- We are now in the fairly unusual situation (from the neuroscientist's viewpoint) of knowing how to show that the model we entertain is wrong without having an explicit expression for this model...
- We need a way to find candidates for the CI: $\lambda(t \mid \mathcal{H}_t)$.

What Do We "Put" in $\mathcal{H}_t$?

- It is common to summarize the stationary discharge of a neuron by its inter-spike interval (ISI) histogram.
- If the latter histogram is not a pure decreasing mono-exponential, that implies that $\lambda(t \mid \mathcal{H}_t)$ will at least depend on the elapsed time since the last spike: $t - t_l$.
- For the real data we saw previously we also expect at least a dependence on the length of the previous inter spike interval, isi_1. We would then have:

$$\lambda(t \mid \mathcal{H}_t) = \lambda(t - t_l, isi_1)$$

What About The Functional Form?

- We haven't even started yet and we are already considering a function of at least 2 variables: $t - t_l, isi_1$. What about its functional form?
- Following Brillinger (1988) we discretize our time axis into bins of size h small enough to have at most 1 spike per bin.
- We are then lead to a binomial regression problem.
- For analytical and computational convenience we are going to use the logistic transform:

$$\log\left(\frac{\lambda(t - t_l, isi_1)\, h}{1 - \lambda(t - t_l, isi_1)\, h}\right) = \eta(t - t_l, isi_1)$$

Smoothing spline

- Since cellular biophysics does not provide much guidance on how to build $\eta(t - t_l, isi_1)$ we have chosen to use the nonparametric smoothing spline approach implemented in the `gss` package.
- $\eta(t - t_l, isi_1)$ is then uniquely decomposed as :

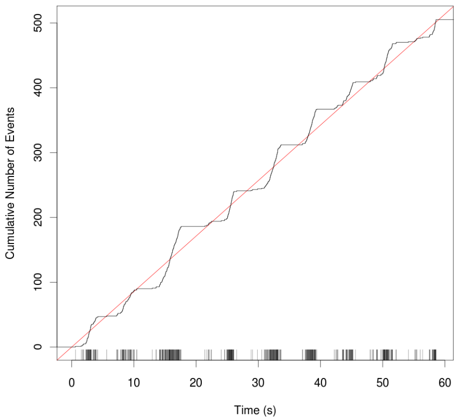
$$\eta(t - t_l, isi_1) = \eta_\emptyset + \eta_l(t_t - l) + \eta_1(isi_1) + \eta_{l,1}(t - t_l, isi_1)$$

- Where for instance:

$$\int \eta_1(u) du = 0$$

the integral being evaluated on the definition domain of the variable isi_1.

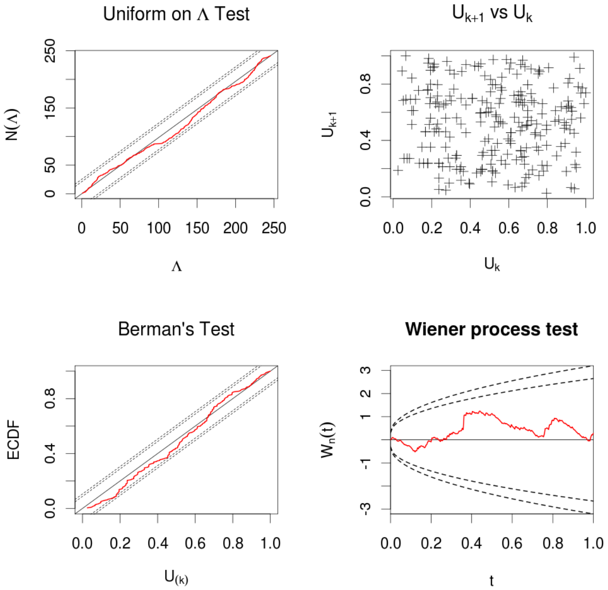
Application to real data



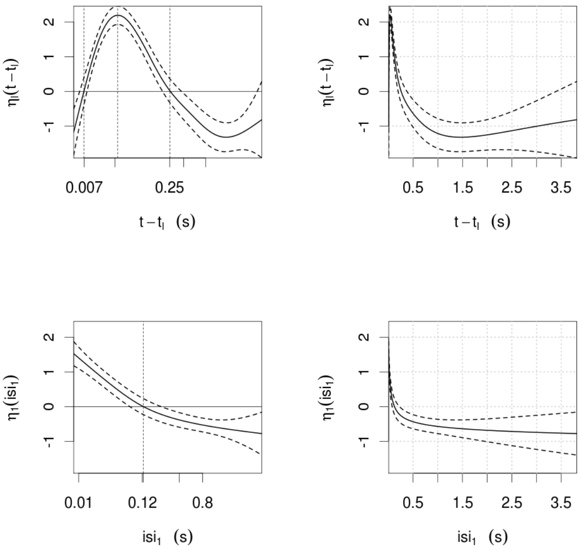
We fitted to the last 30 s of the data set the following additive model:

$$event \sim \sqrt[9]{t - t_l} + \sqrt[10]{isi_1} \, .$$

The tests applied to the first 30 s



The functional forms



Conclusions

- We have now a procedure to fit actual spike trains in a routine fashion.
- We can pass challenging goodness of fit tests.
- The full set of functions required by the analysis we just described is available in the `STAR` (Spike Train Analysis with R) package on `CRAN`.

Acknowledgments

We want to thank:

- Vilmos Prokaj, Olivier Faugeras and Jonathan Touboul for pointing Donsker's theorem to us.
- Carl van Vreeswijk for discussion on the tests.
- The GDR 2904, Systèmes multi-électrodes et traitement du signal appliqués à l'étude réseaux neuronaux, for funding us.