Energy Spectra and Coherence for Point Processes

Rob Dunne, Suzanne Scott, Laurence Wilson, Claus Hallwirth

Tuesday, June 16, 2020

Contents

1 Introduction

We consider the question of analysing a point process (a collection of "times") in the frequency domain. This raises the question of how, if we model a point process as series of spikes at particular times, we transform the time domain to the frequence domain. We mention here that:

- there are many excellent packages in R (R Core Team, 2013), for the time domain analysis of point processes. There appears to be a dearth of software for the frequence analysis of point processes;
- the choice to model the data as a "spike train" may be more natural in some settings than others. The term "spike train" comes from the neuroscience literature. (Uri Eden, 2019; Pouzat and Chaffiol, 2009) model the firing times of neurons as a point process that is then analysed as a "spike train" where each event is an electrical pulse, a signal of zero time duration with some particular mathematical properties.

The frequency domain has several advantages in that subtle structure can be detected which may be difficult to observe in the time domain. Secondly, the time domain estimators can have problems with sensitivity to weak non-stationarity. Jarvis and Mitra (2001) is a key reference in this area. As they comment, most of the literature is targeted at either spectral analysis of continuous processes or at the analysis of point processes but in the time domain.

We could find no package in R that explicitly offered the calculation of an energy spectra for a point process. The Chronux Matlab package (Mitra and Bokil, 2007; Chronux, 2020), offered this so we have taken a small number of functions and coded them in R. We have kept the same names (`mtspectrumpt` and `coherencypt`) for uses familiar with Chronux, but the calling parameters may have changed.

2 point processes, and Dirac δ "functions"

As Brillinger (1994) says "A time series Y is a wiggly line $Y(t), -\infty < t < \infty$. A point process N is a collection of times $\{\tau_j, j = 0, \pm1, \pm2, \ldots\}$. (It will be assumed that the τ_j are distinct.) A marked point process J is a collection of times and associated quantities (marks) $\{(\tau_j, M_j), j = 0, \pm1, \pm2, \ldots\}$."

So say we have a point process. This is just a collection of times and it can be specified in a number of ways as shown in figure 1. $d\bar{N}(t)/dt$ is a series of delta functions.

An electrial pulses is commonly analysed as a Dirac delta. The Dirac delta can be loosely thought of as a discontinuous function on the real line which is zero everywhere except at the origin, where it is undefined (or infinite, if you prefer);

$$\delta(x) = \begin{cases} \text{undefined}, & x = 0 \\ 0, & x \neq 0. \end{cases}$$

However, it has the property that, for some mild conditions on f,

$$\int_{-\infty}^{\infty} f(x)\,\delta(x)dx = f(0).$$

The Dirac delta is a generalized function. No function in the traditional sense has this property.

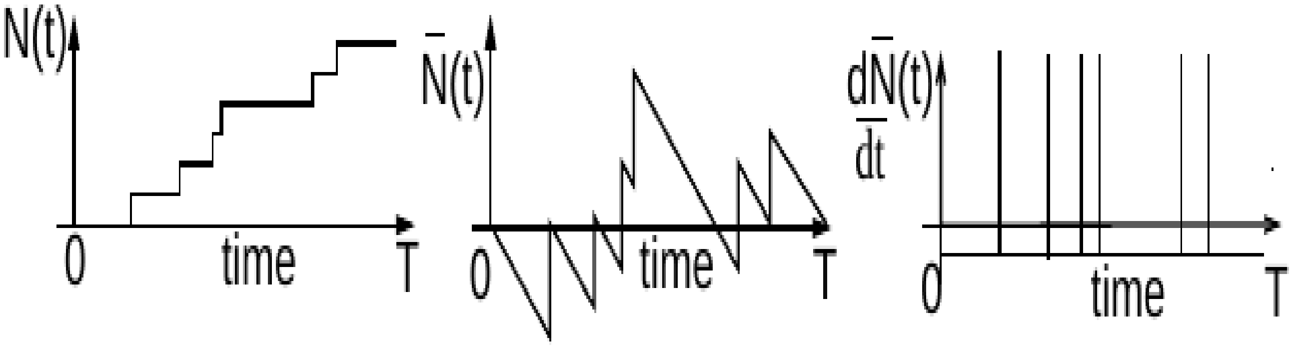

Figure 1: Several specifications for point process data. The counting process $N(t)$ is the number of spikes which occur up to time t, $\bar{N}(t) = N(t) - \lambda t$, where λ is the mean rate. $d\bar{N}(t)/dt$, is a series of delta functions. From Jarvis and Mitra (2001).

3 autocovariance, spectra and coherence

The autocovariance is a function that gives the covariance of the process with itself at pairs of time points. It is

$$\mu(\tau) + \lambda\delta(\tau) = \frac{E[d\bar{N}(t)d\bar{N}(t+\tau)]}{dtd\tau}.$$

The spectrum $S(f)$ is the Fourier transform of the autocovariance function

$$S(f) = \lambda + \int_{\infty}^{\infty} \mu(\tau) \exp(-2\pi i f \tau) d\tau$$

We can extend these definitions to simultaneously recorded spike trains. The cross-covariance is a function that gives the covariance of one process with points on another process

$$\mu_{ab}(\tau) + \lambda_{ab}\delta_{ab}(\tau) = \frac{E[d\bar{N}(t)d\bar{N}(t+\tau)]}{dt d\tau}.$$

The cross-spectrum $S(f)$ is the Fourier transform of the cross-covariance function

$$S_{ab}(f) = \lambda + \int_{\infty}^{\infty} \mu_{ab}(\tau) \exp(-2\pi i f \tau) d\tau$$

and the coherence is

$$\gamma(f) = \frac{S_{ab}(f)}{\sqrt{S_{aa}(f) S_{bb}(f)}}.$$

The spectrum is real and positive but the coherency is complex valued. The modulus of the coherency, which is known as the coherence , can only vary between zero and one. This makes coherence particularly attractive for detecting relationships between spike trains as it is insensitive to the mean spike rates.

4 Spectral Estimators

Taking our data to be $d\bar{N}(t) = x(t)$, say, we define the energy of the signal as

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt.$$

An energy spectral density describes how the energy is distributed with frequency. Parseval's theorem gives us an alternate expression for the energy of the signal:

$$\int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |\hat{x}(f)|^2 \, df,$$

where

$$\hat{x}(f) = \int_{-\infty}^{\infty} e^{-2\pi i f t} x(t) \, dt$$

is the Fourier transform of the signal and f is the frequency in Hz (cycles per second). Since the integral on the right-hand side is the energy of the signal, the integrand $|\hat{x}(f)|^2$ can be interpreted as a density function describing the energy per unit frequency contained in the signal at the frequency f. In light of this, the energy spectral density of a signal $x(t)$ is defined as $S_{xx}(f) = |\hat{x}(f)|^2$. This is the Periodogram.

Jarvis and Mitra (2001) warn that the periodigram is seriously flawed. A improved way to get a spectral estimator is the modulus squared of the Fourier transform of the data multiplied by an envelope function $h(t)$, known as a taper, such that

$$\int_0^T h(t)^2 dt = 1.$$

Then the estimator is

$$
\begin{aligned}
S_{xx} &= \int_0^T h(t) \exp\{-2\pi i f t\} d\bar{N}(t) \\
&= \sum_{j=1}^{N(T)} h(t_j) \exp\{-2\pi i f t_j\} - \frac{N(T)H(f)}{T}
\end{aligned}
$$

where $H(f)$ is the Fourier transform of the taper. $N(T)/T$ is an estimate of the population parameter λ.

The basic idea of multitaper spectral estimation is to average the spectral estimates from several orthogonal tapers.

5 Example 1

Need a simulated example here

```
  library(poisson)
aa1<-hpp.event.times(10, 1000, num.sims = 1, t0 = 0)
S1<-mtspectrumpt(aa, Fs  =   100)
plot(S1$f, S1$S,type="l")

aa2<-hpp.event.times(20, 1000, num.sims = 1, t0 = 0)
S2<-mtspectrumpt(aa2, Fs  =   100)
plot(S2$f, S2$S,type="l")

system.time(C12<-coherencypt(aa1,aa2,Fs  =   500))
plot(C12$f,C12$C12,type="l")

rate <- 100
target <- 1000
intensity <- function(t){ pmin(t/3, 1)}
cc1<-nhpp.event.times(rate, target, intensity)
S3<-mtspectrumpt(cc1, Fs  =   500)
plot(S3$f, S3$S,type="l")
```

Listing 1: Example 2.

6 Example 2

X-linked severe combined immunodeficiency (SCID-X1) is an inherited disease caused by inactivating mutations in the gene encoding the interleukin 2 receptor common gamma chain (IL2RG), which is located on the X-chromosome. SCID-X1 gene therapy clinical trials were conducted in Paris and London. The clinical results differed and the question was raised of the effect of different transduction conditions on transduced cells. Human peripheral blood (PB) CD34+ cells were transduced with γ-retroviral vectors according to the transduction protocols employed in the initial clinical trials. 14 Cells cultured under London trial conditions showed relatively higher retention of CD34 expression at the completion of the transduction period, compared to a loss of CD34-positivity in about half the cells transduced under the Paris trial conditions. Paris conditions, on the other hand, promoted higher levels of both proliferation and transduction, see Hallwirth et al. (2015).

```
data(integrationSites)
Paris<-integrationSites$Paris
```

```
Paris<-as.vector(as.matrix(Paris))
Paris<-Paris/1e+09
London<-integrationSites$London
London<-as.vector(as.matrix(London))
London<-London/1e+09

system.time(S.Paris<-mtspectrumpt(Paris, Fs = 1000))
#   user  system elapsed
#294.788  49.709 210.565
plot(S.Paris$f, S.Paris$S)

system.time(C12<-coherencypt(Paris,London,Fs = 500))
#user  system elapsed
#369.944  81.602 294.154
plot(C12$f,C12$C12,ylim=c(0.8,1),type="l")
```

Listing 2: Example 2.

7 Chronux

Chronux is an open-source software package being developed for the analysis of neural data. It is a collaborative research effort based at the Mitra Lab in Cold Spring Harbor Laboratory that has grown out of the work of several groups. Chronux routines may be employed in the analysis of both point process and continuous data, ranging from preprocessing, exploratory and confirmatory analysis. Chronux is currently implemented as a Matlab toolbox.

The Chronux website at http://chronux.org/ is the central location for information about the current and all previous releases of Chronux. The home page contains links to pages for downloads, people, recent news, tutorials, various files, documentation and our discussion forum. There are also Google groups for Chronux discussion and Chronux announcements which you can join. See also "Observed Brain Dynamics", Partha Mitra and Hemant Bokil, Oxford University Press, New York, 2008.

Chronux is an open source project released under the GNU Public License GPL v2.

References

Baddeley, A. and Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12(1):1–42.

Brillinger, D. R. (1994). Time series, point processes, and hybrids. *Canadian Journal of Statistics*, 22(2):177–206.

Cebrián, A. C., Abaurrea, J., and Asín, J. (2015). NHPoisson: An R Package for Fitting and Validating Nonhomogeneous Poisson Processes. *Journal of Statistical Software*, 64(1):1–25.

Chronux (2020). Chronux analysis software. `http://chronux.org/`. Accessed: Friday, October 9, 2020.

Cohen, E. A. K. (2014). Multi-wavelet coherence for point processes on the real line. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2649–2653. ISSN: 2379-190X.

Hallwirth, C., Garg, G., Peters, T., Kramer, B., Malani, N., Hyman, J., Ruan, X., Ginn, S., Hetherington, N., Veeravali, L., Shahab, A., Ranganathan, S., Chia Wei, L., Liddle, C., Thrasher, A., Bushman, F., Buckley, M., and Alexander, I. (2015). Coherence analysis discriminates between retroviral integration patterns in CD34+ cells transduced under differing clinical trial conditions. *Molecular Therapy - Methods & Clinical Development*, 2.

Harte, D. (2010). PtProcess: An R Package for Modelling Marked Point Processes Indexed by Time. *Journal of Statistical Software*, 35(1):1–32.

Jarvis, M. R. and Mitra, P. P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Comput*, 13(4):717–749.

Mitra, P. and Bokil, H. (2007). *Observed Brain Dynamics*. Oxford University Press.

Peng, R. D. (2007). *ptproc: Multi-dimensional Point Process Models*. R package version 1.5-1.

Pouzat, C. and Chaffiol, A. (2009). Automatic spike train analysis and report generation. An implementation with R, R2html and STAR. *Journal of Neuroscience Methods*, 181(1):119–144.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Uri Eden (2019). *Chapter 2: Introduction to Point Processes*. `http://www.stat.columbia.edu/~liam/teaching/neurostat-fall19/`, [Accessed: Thursday, December 26, 2019].