Spectral estimation of Hawkes count data in discrete time

Felix Cheysson[1,2] Gabriel Lang[1] Laurence Watier[2]

[1]MIA-Paris, UMR 518 AgroParisTech / INRA, Paris, France
[2]B2PHI, UMR 1181 Institut Pasteur / Inserm / University of Versailles St-Quentin-en-Yvelines, Versailles, France

7th Channel Network Conference
July 10th–12th, 2019

Motivation

Study the dynamics of contagious diseases with regards to explanatory variables.

Attributable fraction for contagious diseases

- Common autoregressive models cannot evaluate the repercussion of externalities on self-excitement.
- Potentially rarely occurring events.

$\rightarrow$ Hawkes process (Meyer, Elias, and Höhle, 2012).

Outline

Outline

Point process

Definition: Point process X on $\mathbb{R}$

Random point pattern on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathrm{N}_{\mathbb{R}}, \mathcal{N}_{\mathbb{R}})$$
$$\omega \mapsto N_X = \{t_i\}$$

where $\mathrm{N}_{\mathbb{R}}$ is the set of locally finite subset of $\mathbb{R}$.

Point process

Definition: Point process X on $\mathbb{R}$

Measurable map $N = N_X$:

$$N : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathrm{N}, \mathcal{N})$$
$$\omega \mapsto N^\omega$$

where N is the set of locally finite counting measures on $\mathbb{R}$.

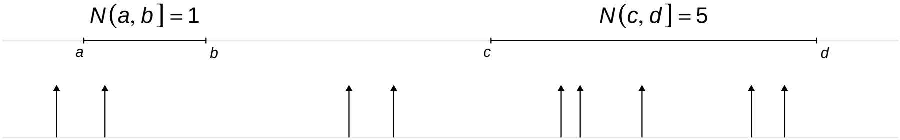
Point process

Definition: Point process X on $\mathbb{R}$

Measurable map $N = N_X$:

$$N : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathrm{N}, \mathcal{N})$$
$$\omega \mapsto N^\omega$$

where N is the set of locally finite counting measures on $\mathbb{R}$.



Hawkes process

Conditional intensity λ^* of point process X

$\lambda^*(t)dt$ is the conditional probability that there will be a point of X between t and $t + dt$, given the realisations of X before t:

$$\lambda^*(t)dt = \mathbb{P}(N(dt) > 0 \mid \{t_j\}, t_j < t)$$

Hawkes process on the real half-line (Hawkes, 1971)

Self-exciting point process defined by its conditional intensity function:

$$\lambda^*(t) = \eta(t) + \sum_{t_j < t} h(t - t_j)$$

where η, h are integrable nonnegative functions and $(t_j)_{j \in \mathbb{N}}$ are realisations of the point process.

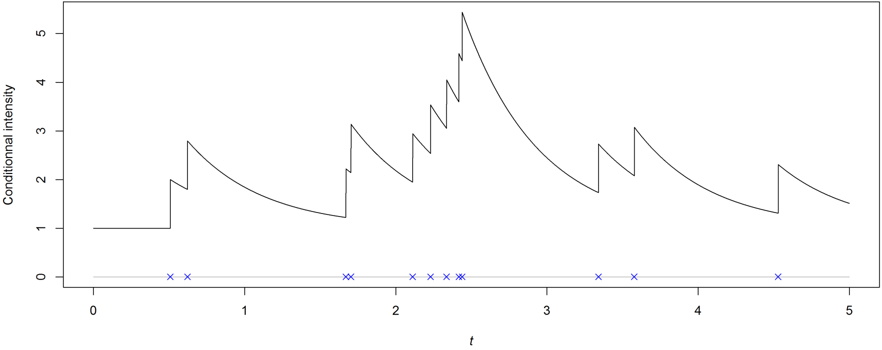
The occurrence of any event increases temporarily the probability of further events occurring.

Hawkes process

Hawkes process on the real half-line

With exponentially decaying intensity:

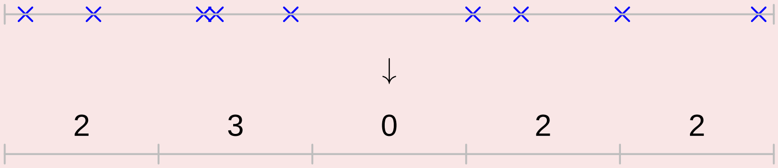
$$\lambda^*(t) = \eta + \sum_{t_j < t} \alpha e^{-\beta(t - t_j)}$$



Outline

Aggregated Hawkes process

Problem: aggregate datasets



Other approaches

- (Kirchner, 2016) Convergence of a well-defined INAR(∞) process to a Hawkes process when the binsize converges to 0.
- (Celeux, Chauveau, and Diebolt, 1995) Convergence of the Stochastic EM algorithm?

Our approach heavily inspired from (Roueff and Von Sachs, 2017): Whittle likelihood for the discrete Hawkes process.

Spectral representation

Bartlett spectrum (Daley and Vere-Jones, 2003, Proposition 8.2.I)

For a second-order stationary point process N on $\mathbb{R}$, then

$$\mathrm{Cov}\left(N(\varphi), N(\psi)\right) = \int_{\mathbb{R}} \widetilde{\varphi}(\xi)\widetilde{\psi^*}(\xi)\Gamma(d\xi)$$

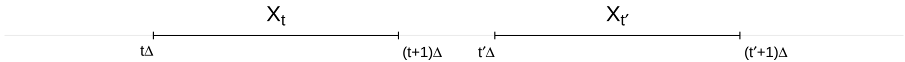
where φ and ψ are functions of rapid decay, $\psi^*(u) = \psi(-u)$, and $\widetilde{\cdot}$ denotes the Fourier transform: $\widetilde{\varphi}(\xi) = \int_{\mathbb{R}} e^{i\xi u}\varphi(u)du$.
The unique measure $\Gamma(\cdot)$ is called the *Bartlett spectrum* of N.

For the stationary Hawkes process N, the Bartlett spectrum admits a density given by (Daley and Vere-Jones, 2003, Example 8.2(e))

$$\gamma(\xi) = \frac{m}{2\pi}\left|1 - H(\xi)\right|^{-2}$$

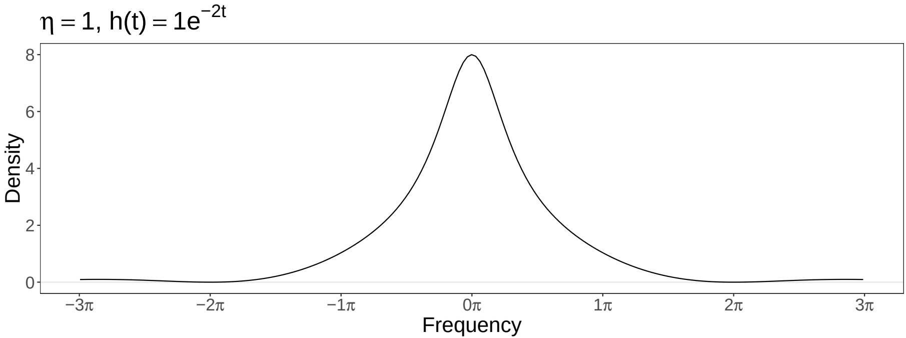
with $m = \mathbb{E}\left[N(0,1]\right] = \frac{\eta}{1-\int_{\mathbb{R}} h(t)dt}$ and $H(\xi) = \int_{\mathbb{R}} e^{i\xi u}h(u)du$.

Spectral representation of the Hawkes process



For the stationary Hawkes process $\{X_t\}_{t\in\mathbb{R}} = \{N(t\Delta, (t+1)\Delta)\}_{t\in\mathbb{R}}$, its spectral density is given by

$$\widehat{\gamma}(\xi) = m\,\Delta\,\mathrm{sinc}^2\left(\frac{\xi}{2}\right)\left|1 - H\left(\frac{\xi}{\Delta}\right)\right|^{-2}$$



The Whittle likelihood (Whittle, 1953)

$$\widehat{\theta}_n = \underset{\theta \in \Theta, \Theta \text{ compact}}{\arg\min} \quad \mathcal{L}_w(\theta)$$

where

$$\mathcal{L}_w(\theta) = \sum_{\omega \in \Omega} \log \widehat{\gamma}(\omega, \theta) + \frac{I_n(\omega)}{\widehat{\gamma}(\omega, \theta)},$$

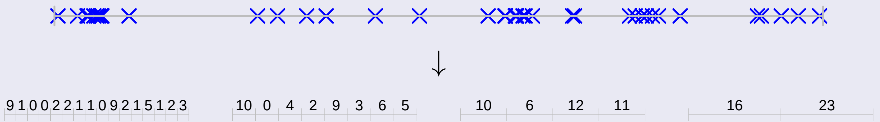
$I_n(\omega)$ is the periodogram of (X_k) and $\Omega = \left(\frac{2k\pi}{n}\right)_{0 \leq k < n}$.

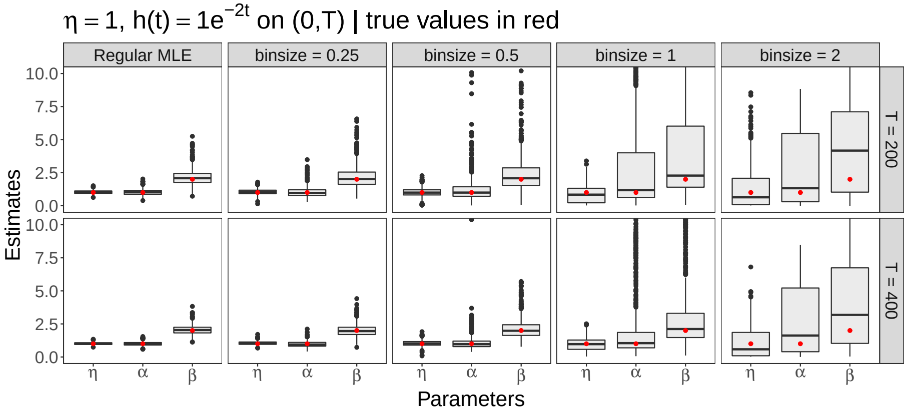
Theorem (Dzhaparidze, 1986)

Given α-mixing conditions on (X_k), $\widehat{\theta}_n$ is consistent and asymptotically normal.

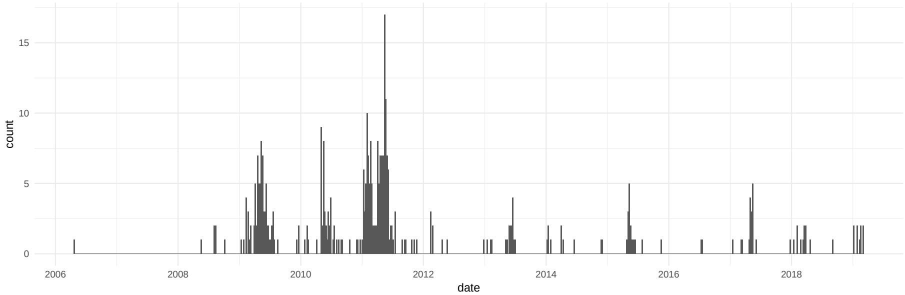
Computationally efficient : $\mathcal{O}(n \log n)$ for the periodogram, then $\mathcal{O}(n)$ for each iteration.

Simulation for the Whittle estimator





Application to a Measles dataset[1]



- Estimated excitation function: $\widehat{h}(t - t_j) = 0.76\, e^{-0.86(t-t_j)}$
- Estimated duration of contagion: $(0.86)^{-1} = 1.16$ weeks $= 8.14$ days

Communicability (Centers for Disease Control and Prevention, 2015)

4 days before to 4 days after rash onset.

[1] Weekly incidence of measles in a French department (Santé publique France)

Perspectives: Extending to non-stationary Hawkes process

Locally stationary Hawkes (Roueff, Sachs, and Sansonnet, 2016)

Given smoothness conditions on $\eta^{<LS>}$ and $h^{<LS>}(\cdot;\cdot)$,

$$N_T(\cdot - Tu) \xrightarrow{\mathcal{D}} N(\cdot; u)$$

where $N(\cdot; u)$ is the stationary Hawkes process at location u.

$\updownarrow$ **?**

Whittle estimator for non stationary time series (Dahlhaus, 1997)

If $X_{t,T}$ has a locally stationary spectral representation (+ *conditions*)

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^0_{t,T}(\lambda) d\xi(\lambda),$$

the Whittle estimator $\widehat{\theta}_T$ is consistent and asymptotically normal.

For Further Reading I

Celeux, Gilles, Didier Chauveau, and Jean Diebolt (1995). *On Stochastic Versions of the EM Algorithm*. Research Report RR-2514. INRIA.

Centers for Disease Control and Prevention (2015). *Epidemiology and Prevention of Vaccine-Preventable Diseases*. Ed. by Jennifer Hamborsky, Andrew Kroger, and Charles (Skip) Wolfe. 13th ed. Washington D.C.: Public Health Foundation.

Dahlhaus, R. (1997). "Fitting time series models to nonstationary processes". In: *Ann. Stat.* 25.1, pp. 1–37. ISSN: 00905364. DOI: 10.1214/aos/1034276620.

Daley, D J and D Vere-Jones (2003). *An Introduction to the Theory of Point Processes*. Vol. I. Probability and its Applications. New York: Springer-Verlag, pp. xviii+573. ISBN: 0-387-95541-0. DOI: 10.1007/b97277. arXiv: arXiv:1011.1669v3.

For Further Reading II

Dassios, Angelos and Hongbiao Zhao (2013). "Exact simulation of Hawkes process with exponentially decaying intensity". In: *Electronic Communications in Probability* 18.62, pp. 1–13. ISSN: 1083-589X. DOI: 10.1214/ECP.v18-2717.

Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. ISBN: 978-1-4612-9325-5. DOI: 10.1007/978-1-4612-4842-2. URL: http://link.springer.com/10.1007/978-1-4612-4842-2.

Hawkes, Alan G (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, p. 83. ISSN: 00063444. DOI: 10.2307/2334319.

Kirchner, Matthias (2016). "Hawkes and INAR(∞) processes". In: *Stochastic Processes and their Applications* 126.8, pp. 2494–2525. ISSN: 03044149. DOI: 10.1016/j.spa.2016.02.008. arXiv: arXiv:1509.02007v1.

For Further Reading III

Meyer, Sebastian, Johannes Elias, and Michael Höhle (2012). "A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence". In: *Biometrics* 68.2, pp. 607–616. ISSN: 0006341X. DOI: 10.1111/j.1541-0420.2011.01684.x. arXiv: 1508.05740.

Møller, Jesper and Jakob G. Rasmussen (2005). "Perfect Simulation of Hawkes Processes". In: *Advances in Applied Probability* 37.3, pp. 629–646.

Ogata, Y. (1981). "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1, pp. 23–31. ISSN: 0018-9448. DOI: 10.1109/TIT.1981.1056305.

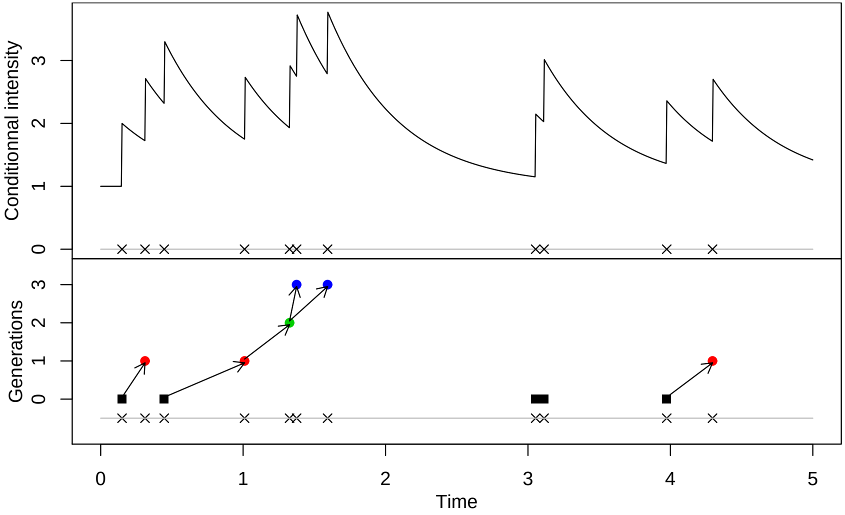
Roueff, F. and R. Von Sachs (Apr. 2017). "Time-frequency analysis of locally stationary Hawkes processes". In: *ArXiv e-prints*. arXiv: 1704.01437 [math.ST].

For Further Reading IV

Roueff, François, Rainer von Sachs, and Laure Sansonnet (2016). "Locally stationary Hawkes processes". In: *Stoch. Process. their Appl.* 126.6, pp. 1710–1743. ISSN: 03044149. DOI: 10.1016/j.spa.2015.12.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304414915003075.

Whittle, P. (1953). "Estimation and information in stationary time series". In: *Arkiv för Matematik* 2.5, pp. 423–434. ISSN: 18712487. DOI: 10.1007/BF02590998.

Hawkes process as a branching process

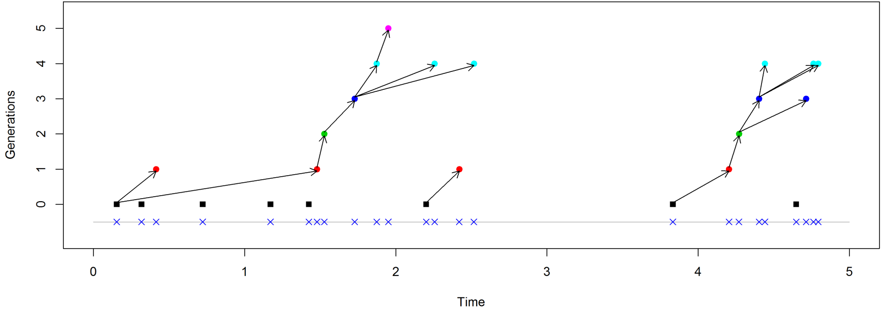


Epidemiological interpretation

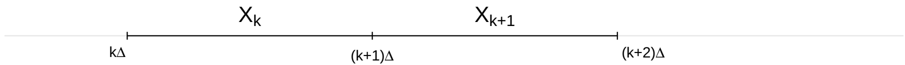
Basic reproduction number

Mean number of infections caused by an individual

$$r = \int_0^\infty h(t)dt$$
$$= \alpha/\beta \qquad \text{for exponentially decaying intensity}$$

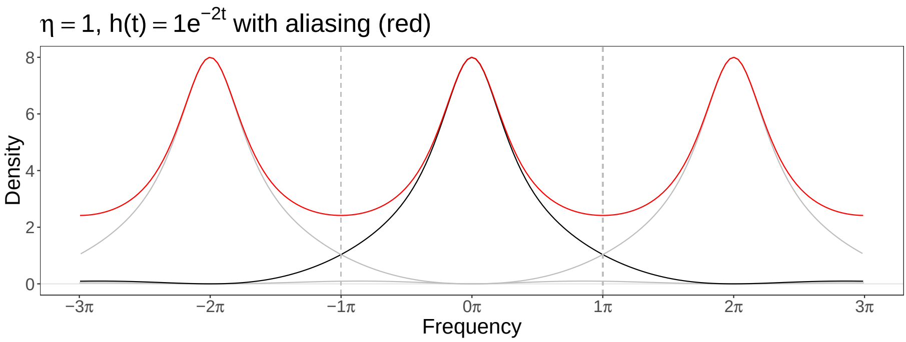


Spectral representation of the discrete Hawkes process

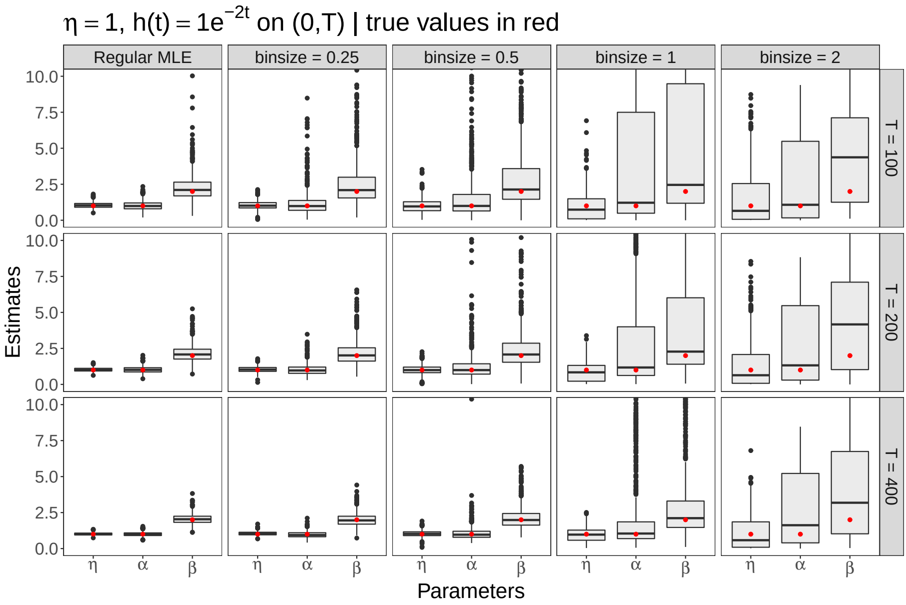


Given a discrete sample $\{X_k\}_{k\in\mathbb{Z}} = \{N(k\Delta,(k+1)\Delta)\}_{k\in\mathbb{Z}}$ with a sampling step of 1, its spectral density is given by

$$\widehat{\gamma}_1(\xi) = \sum_{k\in\mathbb{Z}} \widehat{\gamma}(\xi + 2k\pi)$$

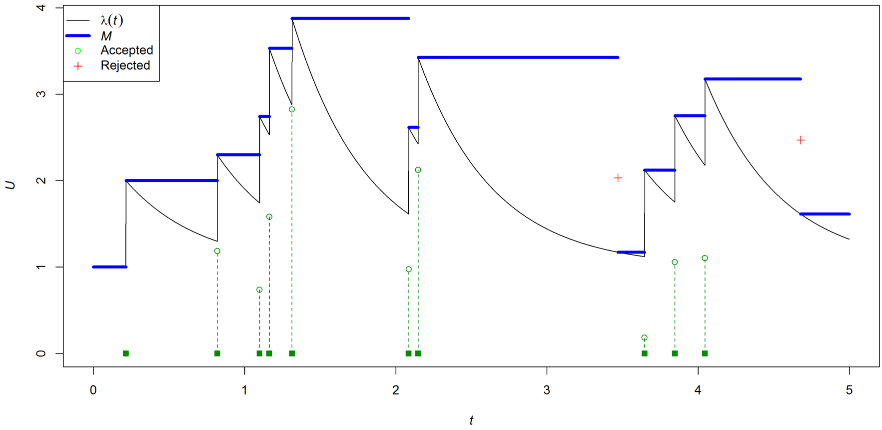


Simulation for the Whittle estimator



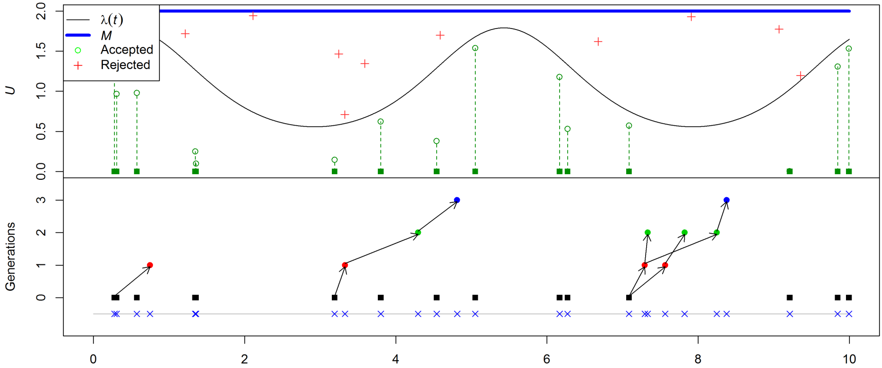
Simulate Hawkes in *R* (Ogata, 1981)

```
sim <- hawkes(T=10, lambda=1, alpha=1, beta=2)
plot(sim)
```



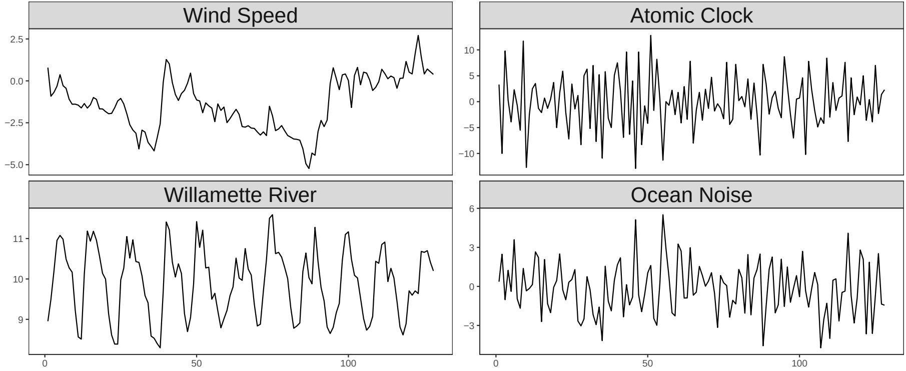
Simulate Hawkes with inhomogeneous background intensity in *R* (Møller and Rasmussen, 2005; Dassios and Zhao, 2013)

```
int <- function(t) exp(.5*cos(2*pi*t/5)+.3*sin(2*pi*t/5))
sim <- twinstim(T=10, fun=int, M=2, alpha=1, beta=2)
plot(sim$immigrants)
plot(sim)
```



Introduction to spectral analysis

We consider four examples of time series $x_1, x_2, \ldots, x_{128}$. How would you describe them?

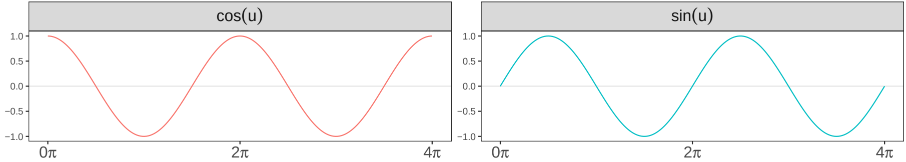


Spectral analysis describes x_t's by comparing them to sines and cosines.

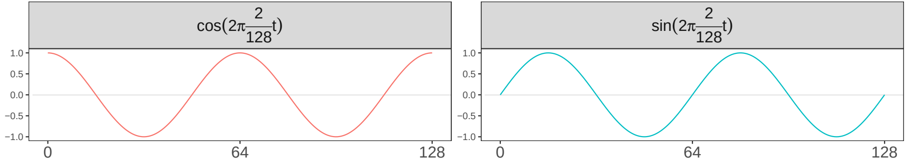
Sines and cosines?

The functions $\sin$ and $\cos$ are 2π-periodic: for $u \in \mathbb{R}$ and $k \in \mathbb{Z}$,

$$\cos(u + 2k\pi) = \cos(u)$$



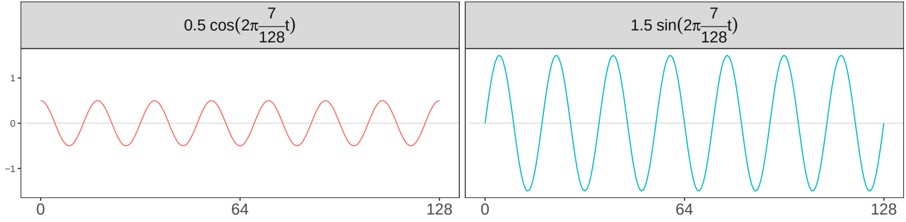
Let $u = 2\pi \frac{2}{128} t$ for $t = 1, 2, \ldots, 128$.



$\frac{2}{128}$ can be interpreted as 2 cycles over the time span of 128.

Sines and cosines?

Similarly, with $u = 2\pi\frac{7}{128}t$ for $t = 1, 2, \ldots, 128$.



- $\cos(2\pi\frac{k}{n}t)$ and $\sin(2\pi\frac{k}{n}t)$ have k cycles per n time steps.
- The quantity $f = \frac{k}{n}$ is called the *frequency* of the sine or cosine.
  - It is the amount (or rather fraction) of cycles per time step.
  - If f is small (large), the sine is said to have low (high) frequency.
- The *period* $T = \frac{1}{f}$ is the time steps needed for a full cycle.
- The *amplitude* is the maximum range of variation and is equal to 1 for the functions $\sin$ and $\cos$.

Sines and cosines?

Summing up sines and cosines of different amplitudes and frequencies create time series that resemble actual data.

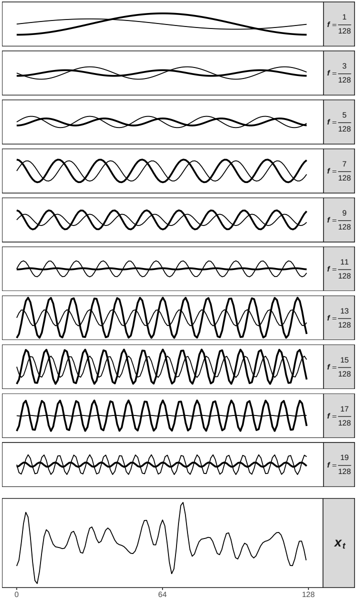
Goal of spectral analysis

Given a time series x_t, figure out its *Fourier representation*, *i.e.* its decomposition into sines and cosines:

$$x_t = \sum_k a_k \cos\left(2\pi \frac{k}{n} t\right) + b_k \sin\left(2\pi \frac{k}{n} t\right)$$

Actually easy to compute a_k and b_k:

$$a_k \propto \operatorname{Cov}\left\{x_t, \cos\left(2\pi \frac{k}{n} t\right)\right\}$$



The spectrum and periodogram

$\sigma_k^2 = a_k^2 + b_k^2$, the squared amplitude of the sine-cosine pair, highlights the importance of the frequency $\frac{k}{n}$ in the decomposition of x_t.

- $(\sigma_k^2)_{k=1,\dots,n}$ is called the *spectrum* of the time series x_t.
- If σ_k^2 is large, there are strong patterns of frequency $\frac{k}{n}$.
- The sample spectrum, noted I_k, is called the *periodogram* of x_t.

Decomposition of variance

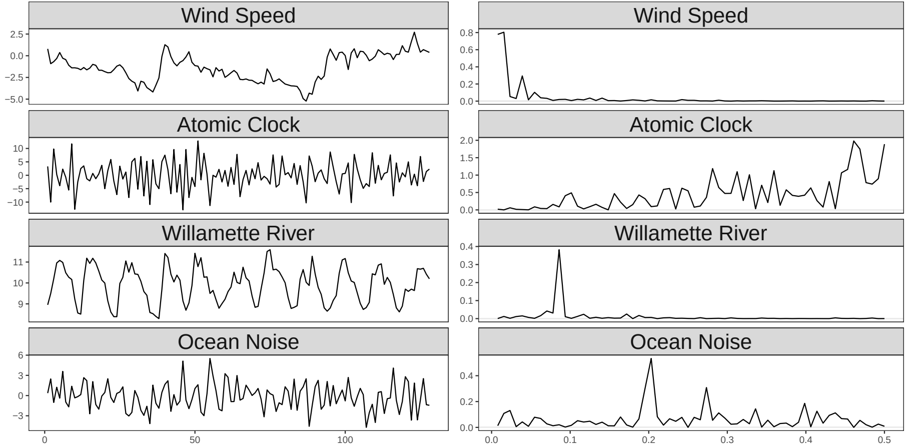
Since the sines and cosines of different frequencies are uncorrelated, then

$$\mathrm{Var}\ x_t = \sum_k (a_k^2 + b_k^2) = \sum_k \sigma_k^2$$

The spectrum (σ_k^2) is the decomposition of the variance of the time series x_t into its different frequencies $\frac{k}{n}$.

Examples

Recall the four examples: here are their periodograms.



An insight into the continuous case

Spectral density

For a second-order stationary process x_t with absolute summable autocovariance $\gamma(h)$, then

$$\gamma(h) = \frac{1}{2\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i \xi h} f(\xi) d\xi$$

where f is called the *spectral density* of x_t.

The spectral density f is the continuous equivalent of the spectrum (σ_k^2) and is the *Fourier transform* of $\gamma(h)$:

$$f(\xi) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{2\pi i \xi h}$$

Locally stationary time series (Dahlhaus, 1997)

$X_{t,T}$ is called locally stationary with transfer function A^0 and trend μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^0_{t,T}(\lambda) d\xi(\lambda),$$

where

(i) $\xi(\lambda)$ is an orthogonal-increments stochastic process with bounded cumulants, and

(ii) there exists a continuous 2π-periodic function $A : [0,1] \times \mathbb{R} \to \mathbb{C}$ s.t.

$$\sup_{t,\lambda} \left| A^0_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| = \mathcal{O}(T^{-1}).$$

Whittle likelihood (cont'd) (Dahlhaus, 1997)

Given a number of regularity assumptions,

$$\widehat{\theta}_n = \underset{\theta \in \Theta, \Theta \text{ compact}}{\arg\min} \quad \mathcal{L}_w(\theta)$$

is consistent and asymptotically normal, where

$$\mathcal{L}_w(\theta) = \frac{1}{4\pi} \sum_{t=1}^{T} \sum_{\omega \in \Omega} \left\{ \log 4\pi^2 f_\theta\left(\frac{t}{T}, \omega\right) + \frac{\widetilde{I}_T(t/T, \omega)}{f_\theta(t/T, \omega)} \right\},$$

$f_\theta(u, \cdot)$ is the local spectrum at location $t = uT$ and

$$\widetilde{I}_T(u, \omega) = \frac{1}{2\pi} \sum_{k: 1 \leq [uT + 0.5 \pm k/2 \leq T]} X_{[uT+0.5+k/2]} X_{[uT+0.5-k/2]} \exp(-i\omega k)$$

is the preperiodogram of X_t, a local periodogram analog at location uT.