Estimation of Hawkes processes from binned observations using Whittle likelihood

Felix Cheysson   Gabriel Lang

Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris.

EcoDep Seminary
November 25^{th} 2020

Motivation

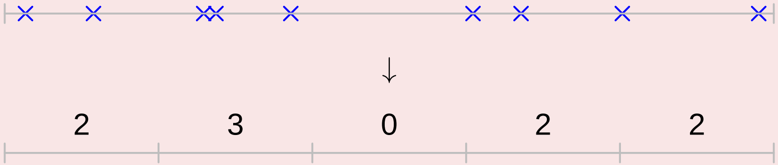
Study the dynamics of contagious diseases and their transmission with respect to risk factors.

Attributable fraction for contagious diseases

- Autoregressive models may be difficult to interpret in an epidemiological context.
- Potentially rarely occurring events.

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

Problem: aggregate datasets



Outline

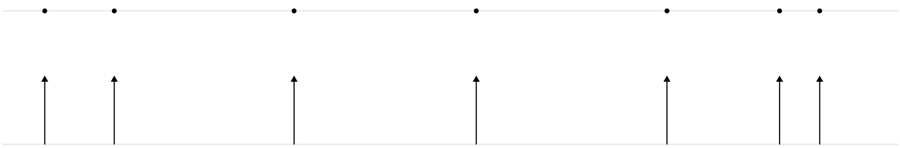
Outline

Point process

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

Point process

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



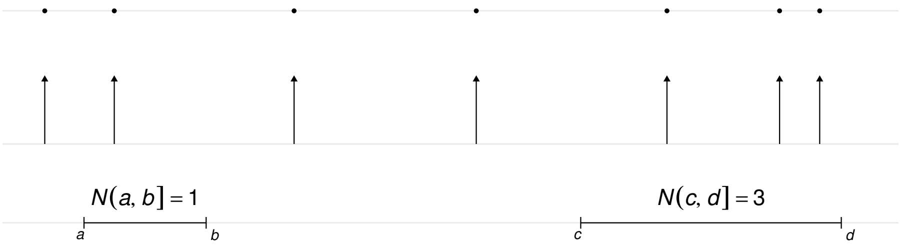
Point process

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

Measurable map N:

$$\begin{aligned} N : (\Omega, \mathcal{F}, \mathbb{P}) &\to (\mathfrak{N}, \mathcal{N}) \\ \omega &\mapsto N(\omega, \cdot) \end{aligned}$$

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



Hawkes process

Conditional intensity λ^* of point process N

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

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

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

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

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

where η, h are integrable nonnegative functions such that $\int h < 1$, and $(t_j)_{j \in \mathbb{N}}$ are realisations of the point process.

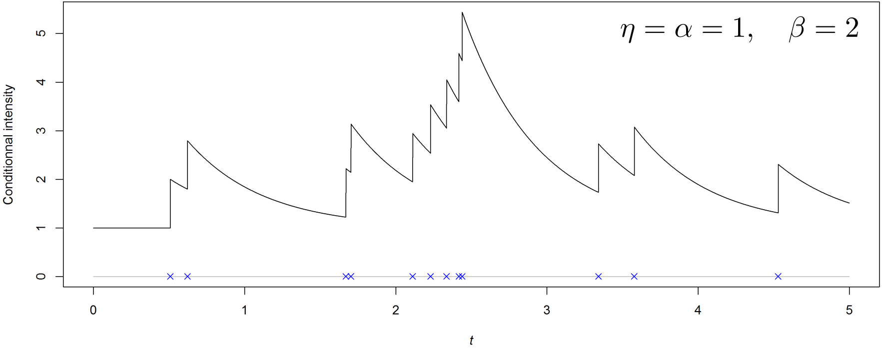
The occurrence of any event increases temporarily the probability of further events occurring.

Hawkes process

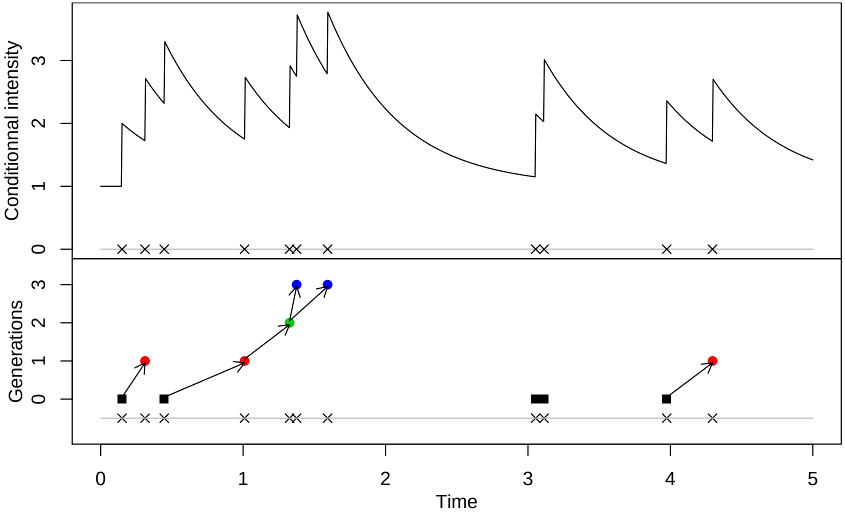
Linear Hawkes process on the real half-line

With exponentially decaying intensity:

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



Hawkes process as a branching process



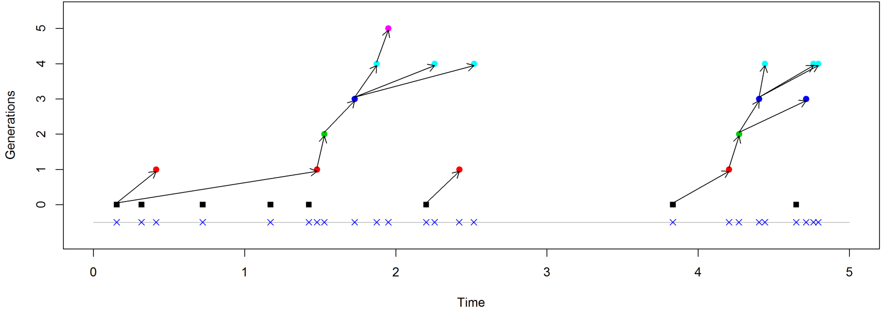
Epidemiological interpretation

Basic reproduction number

Mean number of infections caused by an individual

$$\mu = \int_0^\infty h(t)dt$$

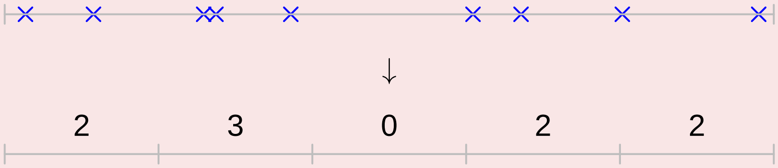
$$= \alpha/\beta \qquad \text{for exponentially decaying intensity}$$



Outline

Motivation

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?

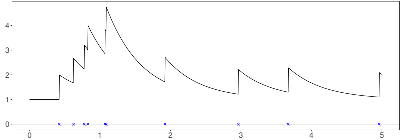
Our approach inspired from (Adamopoulos, 1976; Roueff and Sachs, 2019): Whittle likelihood for Hawkes bin-count sequences.

Idea: spectral approach

Objective: Estimate $\theta = (\eta, h)$ from the count process

Idea: spectral approach

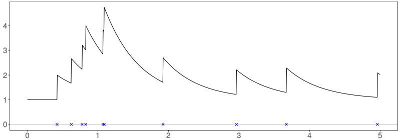
Hawkes process with parameter $\theta = (\eta, h)$



Objective: Estimate $\theta = (\eta, h)$ from the count process

Idea: spectral approach

Hawkes process with parameter $\theta = (\eta, h)$



|

Likelihood of the count process is not tractable

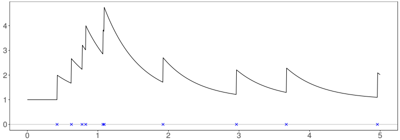
↓

Objective: Estimate $\theta = (\eta, h)$ from the count process

Idea: spectral approach

Time domain

Hawkes process with parameter $\theta = (\eta, h)$



$\big|$

Likelihood of the count process is not tractable

$\downarrow$

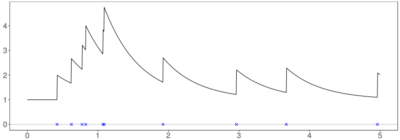
Objective: Estimate $\theta = (\eta, h)$ from the count process

Frequency domain

Idea: spectral approach

Time domain

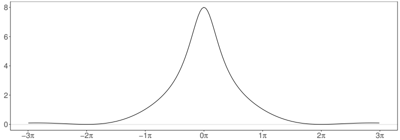
Hawkes process with parameter $\theta = (\eta, h)$



→

Frequency domain

Bartlett spectrum (Daley and Vere-Jones, 2003, Section 8.2)



|

Likelihood of the count process is not tractable

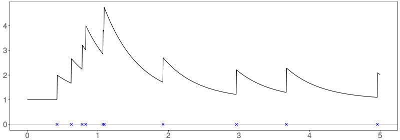
↓

Objective: Estimate $\theta = (\eta, h)$ from the count process

Idea: spectral approach

Time domain

Hawkes process with parameter $\theta = (\eta, h)$



→

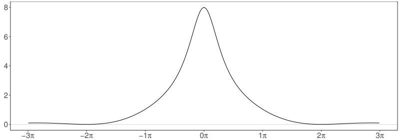
Likelihood of the count process is not tractable

↓

Objective: Estimate $\theta = (\eta, h)$ from the count process

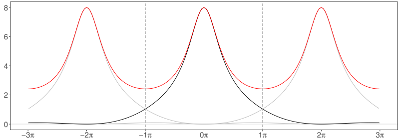
Frequency domain

Bartlett spectrum (Daley and Vere-Jones, 2003, Section 8.2)



Spectral aliasing

↓



Spectral density function: f_θ

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

$$\operatorname{Cov}(N(\varphi), N(\psi)) = \int_{\mathbb{R}} \widetilde{\varphi}(\omega)\widetilde{\psi^*}(\omega)\Gamma(d\omega)$$

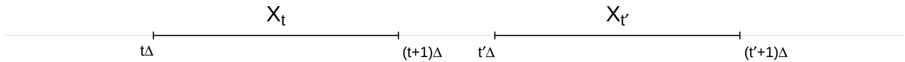
where φ and ψ are functions of rapid decay, $\psi^*(u) = \psi(-u)$, and $\widetilde{\cdot}$ denotes the Fourier transform: $\widetilde{\varphi}(\omega) = \int_{\mathbb{R}} e^{i\omega u}\varphi(u)du$.
The unique measure $\Gamma(\cdot)$ is called the *Bartlett spectrum* of N.

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

$$\gamma(\omega) = \frac{m}{2\pi}|1 - \widetilde{h}(\omega)|^{-2}$$

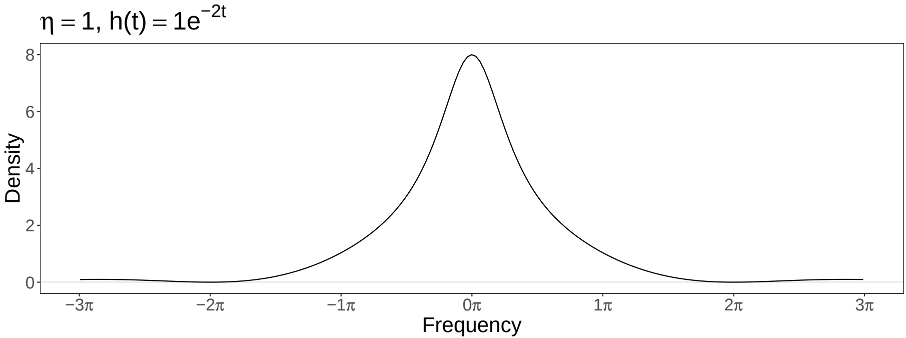
with $m = \mathbb{E}[N(0,1]] = \eta(1-\mu)^{-1}$.

Spectral representation of the bin-count process

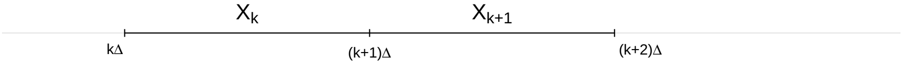


For the Hawkes bin-count process $\{X_t\}_{t\in\mathbb{R}} = \{N(t\Delta, (t+1)\Delta)\}_{t\in\mathbb{R}}$, the spectral density is given by

$$f_{X_t}(\omega) = m\,\Delta\,\mathrm{sinc}^2\left(\frac{\omega}{2}\right)\left|1 - \widetilde{h}\left(\frac{\omega}{\Delta}\right)\right|^{-2}$$

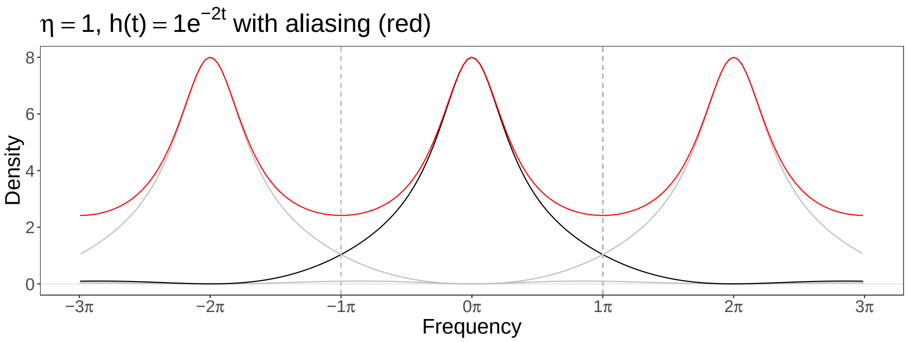


Spectral representation of the bin-count sequence



For the bin-count sequence $\{X_k\}_{k \in \mathbb{Z}} = \{N(k\Delta, (k+1)\Delta)\}_{k \in \mathbb{Z}}$, the spectral density is given by

$$f_{X_k}(\omega) = \sum_{k \in \mathbb{Z}} f_{X_t}(\omega + 2k\pi)$$



The Whittle likelihood (Whittle, 1952)

Consider a bin-count sequence (X_k) with spectral density f_θ. Define

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\arg\min}\ \mathcal{L}_n(\theta)$$

where $\mathcal{L}_n(\theta)$ is the log-spectral likelihood of the process

$$\mathcal{L}_n(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log f_\theta(\omega) + \frac{I_n(\omega)}{f_\theta(\omega)} \right) \mathrm{d}\omega,$$

$I_n(\omega)$ is the periodogram of (X_k).

Asymptotic properties for $\widehat{\theta}_n$

- For Gaussian* processes (Whittle, 1953);
- For linear processes (Hosoya, 1974; Dzhaparidze, 1974);
- For strongly mixing processes (Dzhaparidze, 1986).

Outline

Weak dependence for time series and random fields

- Rosenblatt's strong-mixing coefficient (1956), to measure the dependence between σ-algebras:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup\left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}, B \in \mathcal{B} \right\}.$$

- Strong mixing coefficient for a time series $(X_k)_{k \in \mathbb{Z}}$:

$$\alpha_X(r) := \sup_{k \in \mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{k}, \mathcal{F}_{k+r}^{\infty}\right), \qquad \text{where } \mathcal{F}_a^b = \sigma\left(X_k, a \leq k \leq b\right).$$

- Provides strong moment inequalities (Doukhan, 1994; Rio, 2017), *provided the coefficient decreases fast enough*.
- Other existing mixing coefficients, notably the absolute regularity mixing coefficients.
  - Easily computed for (functions of) Markov processes.

See (Bradley, 2005) for a short review of mixing conditions for time series.

Mixing properties for point processes

Define, for a Borel set $A \in \mathbb{R}$, the cylindrical σ-algebra generated by a point process N on A:

$$\mathcal{E}(A) := \sigma\big(\{N \in \mathfrak{N} : N(B) = m\}, B \in \mathcal{B}(A), m \in \mathbb{N}\big).$$

Strong mixing coefficient for a point process N (Westcott, 1972)

Dependence between past and future events:

$$\alpha_N(r) := \sup_{t \in \mathbb{R}} \alpha\big(\mathcal{E}_{-\infty}^{t}, \mathcal{E}_{t+r}^{\infty}\big), \qquad \text{where } \mathcal{E}_a^b = \mathcal{E}\big((a,b]\big).$$

- Poisson cluster processes are mixing (Westcott, 1971).
- Recent applications of mixing coefficients for point processes (Heinrich and Pawlas, 2013; Poinas, Delyon, and Lavancier, 2019).

Strong mixing properties for Hawkes processes

Theorem

Let N be a stationary Hawkes process on $\mathbb{R}$ with reproduction function $h = \mu h^*$, $\mu = \int_{\mathbb{R}} h$. Suppose that there exists a $\delta > 0$ such that the reproduction kernel h^* has a finite moment of order $1 + \delta$:

$$\nu_{1+\delta} := \int_{\mathbb{R}} t^{1+\delta} h^*(t) \mathrm{d}t < \infty.$$

Then N is strongly mixing and

$$\alpha_N(r) = \mathcal{O}\left(r^{-\delta}\right).$$

Ideas of the proof

We need to bound

$$\alpha_N(r) \coloneqq \sup_{t\in\mathbb{R}} \alpha\big(\mathcal{E}^t_{-\infty}, \mathcal{E}^\infty_{t+r}\big) = \sup_{t\in\mathbb{R}} \sup_{\substack{\mathcal{A}\in\mathcal{E}^t_{-\infty}\\ \mathcal{B}\in\mathcal{E}^\infty_{t+r}}} \Big|\mathrm{Cov}\big(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\big)\Big|,$$

where $\mathbb{1}_{\mathcal{A}}(N)$ is the indicator function of the cylinder set $\mathcal{A}$, *i.e.* for an elementary cylinder set $\mathcal{A}_{B,m} = \{N \in \mathfrak{N} : N(B) = m\}$,

$$\mathbb{1}_{\mathcal{A}_{B,m}}(N) = \begin{cases} 1 & \text{if } N(B) = m, \\ 0 & \text{otherwise.} \end{cases}$$

Ideas of the proof (cont'd)

$$\alpha_N(r) = \sup_{t\in\mathbb{R}} \sup_{\substack{\mathcal{A}\in\mathcal{E}_{-\infty}^{t} \\ \mathcal{B}\in\mathcal{E}_{t+r}^{\infty}}} \left|\mathrm{Cov}\left(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\right)\right| \quad (1)$$

1. Control (1) by the covariance of counts.
   - Hawkes processes are infinitely divisible (Gao and Zhu, 2018, Section 2.1, key property (e));
   - They are positively associated (Evans, 1990, Theorem 1.1);
   - Use of Theorem 2.5 from (Poinas, Delyon, and Lavancier, 2019).
2. Rescale to a single branching process by conditioning on the immigrant process.
3. Control the covariance of counts of a single branching process.
   - Almost sure extinction of the subcritical Galton-Watson tree;
   - Finite moments for the reproduction kernel.
4. Integrate back with respect to the immigrant process.

Asymptotic properties of the Whittle estimator

Direct application of (Dzhaparidze, 1986, Theorem II.7.1).

Consistency

Let $(X_k)_{k\in\mathbb{Z}} = \big(N(k, k+1]\big)_{k\in\mathbb{Z}}$ be the bin-count sequences of a stationary Hawkes process, with spectral density function f_θ. Assume the following regularity conditions on f_θ:

(A1) The true value θ_0 belongs to a compact set $\Theta \subset \mathbb{R}^p$.

(A2) For all $\theta_1 \neq \theta_2$ in Θ, then $f_{\theta_1} \neq f_{\theta_2}$ for almost all ω.

(A3) The function f_θ^{-1} is differentiable with respect to θ and its derivatives $(\partial/\partial\theta_k) f_\theta^{-1}$ are continuous in $\theta \in \Theta$ and $-\pi \leq \omega \leq \pi$.

Further assume that there exists a $\delta > 0$ such that the reproduction kernel h^* has a finite moment of order $2+\delta$. Then the estimator $\widehat{\theta}_n$ is consistent, *i.e.* $\widehat{\theta}_n \to \theta_0$ in probability.

Asymptotic properties (cont'd)

Direct application of (Dzhaparidze, 1986, Theorem II.7.2).

Asymptotic normality

Let $(X_k)_{k\in\mathbb{Z}} = \big(N(k, k+1]\big)_{k\in\mathbb{Z}}$ be the bin-count sequences of a stationary Hawkes process, with spectral density function f_θ. Assume conditions (A1), (A2), (A3) and:

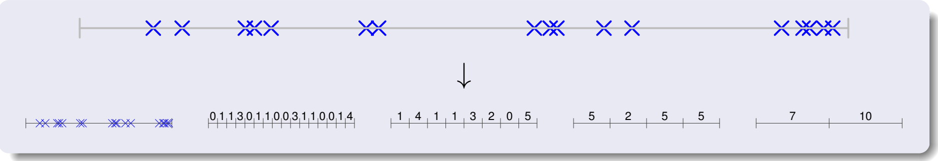
(A4) The function f_θ is twice differentiable with respect to θ and its second derivatives $(\partial^2/\partial\theta_k\partial\theta_l)f_\theta$ are continuous in $\theta \in \Theta$ and $-\pi \leq \omega \leq \pi$.

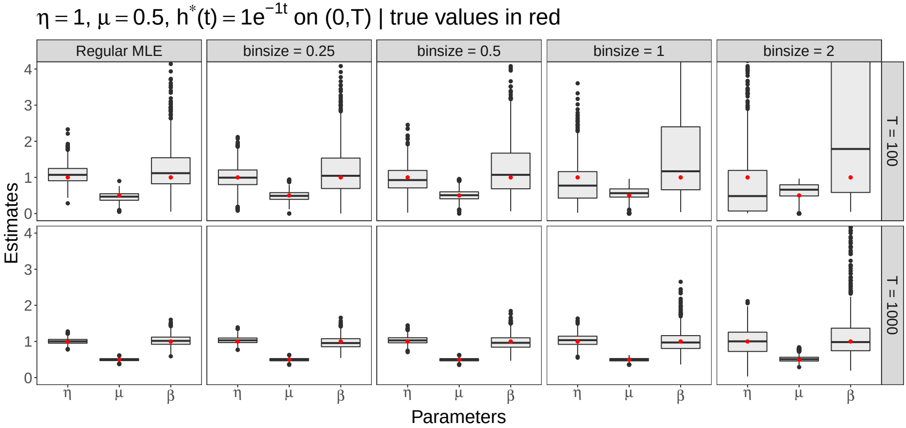
Then the estimator $\widehat{\theta}_n$ is asymptotically normal and

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \underset{n\to\infty}{\sim} \mathcal{N}\left(0, \Gamma_{\theta_0}^{-1} + \Gamma_{\theta_0}^{-1} C_{4,\theta_0} \Gamma_{\theta_0}^{-1}\right).$$

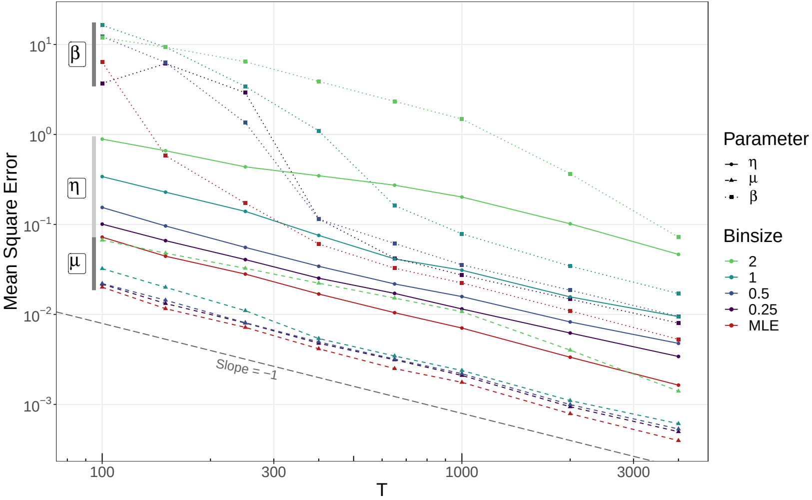
Outline

Simulation for the Whittle estimator

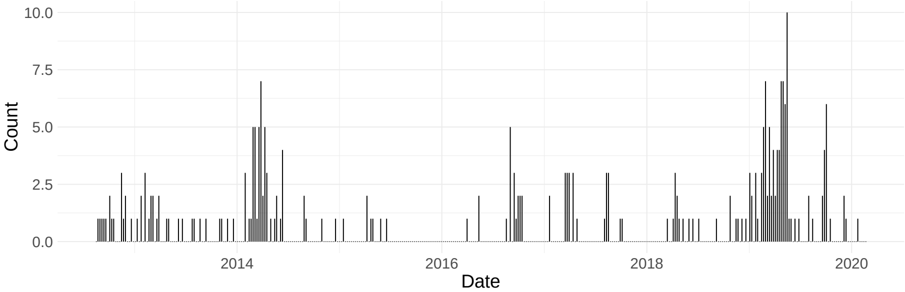




Simulation for the Whittle estimator



Case-study: transmission of Measles in Tokyo[1]



Gaussian reproduction kernel: $h^*(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\nu)^2}{2\sigma^2}\right)$

- $\widehat{\nu} = 9.8 \text{ days}, \widehat{\sigma} = 5.9 \text{ days}$

Epidemiology (Centers for Disease Control and Prevention, 2015)

Incubation period: 10-12 days after exposure.

Transmission period: 4 days before to 4 days after rash onset.

[1] https://www.niid.go.jp/niid/en/survaillance-data-table-english.html

Outline

Conclusion

- Good asymptotic properties, similar to maximum likelihood estimation;
- Easy to implement and flexible, only need to input $\widetilde{h}$;
- Computationally efficient : $\mathcal{O}(n \log n)$ instead of $\mathcal{O}(n^2)$ for maximum likelihood.

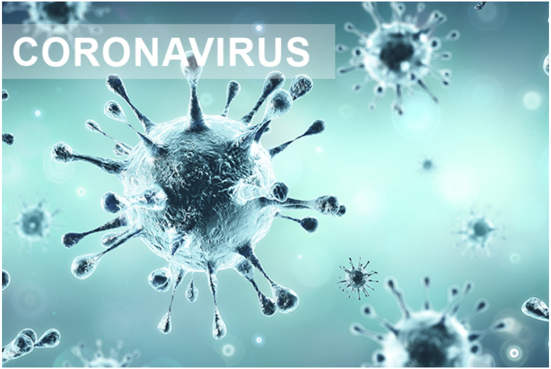
Simulation and estimation methods implemented in R package *hawkesbow*.[2]

Extensions

- **Non causal Hawkes processes**.
- **Multivariate Hawkes processes** (work in progress with Ousmane Boly, Thi Hien Nguyen and Paul Doukhan):
  - Exponential inequalities for multitype Galton-Watson trees.
  - Multivariate Bartlett spectrum (Daley and Vere-Jones, 2003, Example 8.3(c));

[2] https://github.com/fcheysson/hawkesbow

Perspectives



- Non-stationary Hawkes processes: allow all parameters to vary with time and be dependent on explanatory variables;
- Transient explosivity: allow μ to temporarily be higher than 1.
- Spatial dimension: irregular partitioning of areal zones.

For Further Reading I

Adamopoulos, L. (1976). "Cluster models for earthquakes: Regional comparisons". In: *J. Int. Assoc. Math. Geol.* 8.4, pp. 463–475. ISSN: 0020-5958. DOI: 10.1007/BF01028982.

Bradley, Richard C. (2005). "Basic properties of strong mixing conditions. A survey and some open questions". In: *Probab. Surv.* 2.1, pp. 107–144. ISSN: 15495787. DOI: 10.1214/154957805100000104.

Celeux, Gilles, Didier Chauveau, and Jean Diebolt (1995). *On Stochastic Versions of the EM Algorithm*. Tech. rep. RR-2514. INRIA, p. 22.

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.

For Further Reading II

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

– (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Springer. ISBN: 9780387213378.

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

Doukhan, Paul (1994). *Mixing: Properties and Examples*. Springer-Verlag, New York. ISBN: 978-1-4612-2642-0.

Dzhaparidze, K. O. (1974). "A New Method for Estimating Spectral Parameters of a Stationary Regular Time Series". In: *Theory Probab. Its Appl.* 19.1, pp. 122–132. ISSN: 0040-585X. DOI: 10.1137/1119009.

For Further Reading III

Dzhaparidze, Kacha (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-1-4612-9325-5. DOI: 10.1007/978-1-4612-4842-2. arXiv: arXiv:1011.1669v3.

Evans, Steven N. (1990). "Association and random measures". In: *Probab. Theory Relat. Fields* 86.1, pp. 1–19. ISSN: 01788051. DOI: 10.1007/BF01207510.

Gao, Xuefeng and Lingjiong Zhu (2018). "Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues". In: *Queueing Syst.* 90.1-2, pp. 161–206. ISSN: 15729443. DOI: 10.1007/s11134-018-9570-5. arXiv: arXiv:1607.06624v4.

Hawkes, Alan G (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, pp. 83–90. ISSN: 00063444. DOI: 10.2307/2334319.

For Further Reading IV

Heinrich, Lothar and Zbyněk Pawlas (2013). "Absolute regularity and Brillinger-mixing of stationary point processes". In: *Lith. Math. J.* 53.3, pp. 293–310. ISSN: 15738825. DOI: 10.1007/s10986-013-9209-5.

Hosoya, Yuzo (1974). "Estimation problems on stationary time series models". Ph.D. dissertation. Yale University.

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

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: *Adv. Appl. Probab.* 37.3, pp. 629–646.

For Further Reading V

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

Poinas, Arnaud, Bernard Delyon, and Frédéric Lavancier (2019). "Mixing properties and central limit theorem for associated point processes". In: *Bernoulli* 25.3, pp. 1724–1754. ISSN: 1350-7265. DOI: 10.3150/18-BEJ1033. arXiv: 1705.02276.

Rio, Emmanuel (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Vol. 80. Probability Theory and Stochastic Modelling. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-662-54322-1. DOI: 10.1007/978-3-662-54323-8.

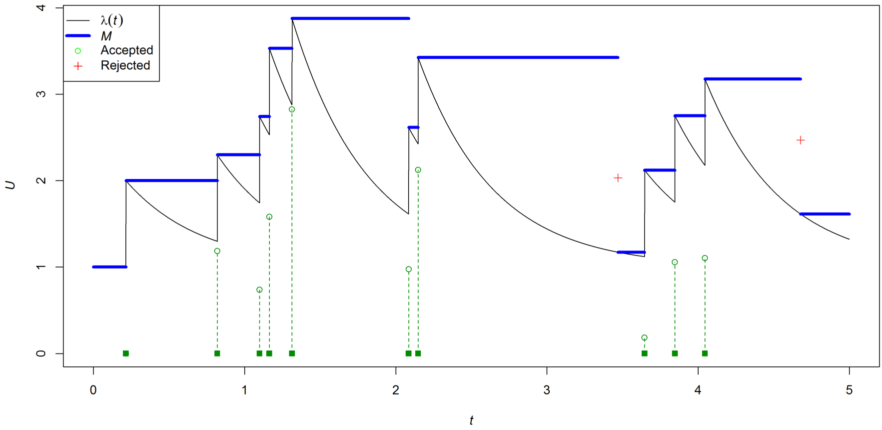
Rosenblatt, M. (1956). "A Central Limit Theorem and a Strong Mixing Condition". In: *Proc. Natl. Acad. Sci.* 42.1, pp. 43–47. ISSN: 0027-8424. DOI: 10.1073/pnas.42.1.43.

For Further Reading VI

- Roueff, François and Rainer von Sachs (2019). "Time-frequency analysis of locally stationary Hawkes processes". In: *Bernoulli* 25.2, pp. 1355–1385. ISSN: 1350-7265. DOI: 10.3150/18-BEJ1023. arXiv: 1704.01437.
- Westcott, M. (1971). "On Existence and Mixing Results for Cluster Point Processes". In: *J. R. Stat. Soc. Ser. B* 33.2, pp. 290–300. DOI: 10.1111/j.2517-6161.1971.tb00880.x.
- – (1972). "The probability generating functional". In: *J. Aust. Math. Soc.* 14.4, pp. 448–466. ISSN: 14468107. DOI: 10.1017/S1446788700011095.
- Whittle, P. (1953). "Estimation and information in stationary time series". In: *Ark. för Mat.* 2.5, pp. 423–434. ISSN: 0004-2080. DOI: 10.1007/BF02590998.
- Whittle, Peter (1952). "Some results in time series analysis". In: *Scand. Actuar. J.* 1952.1-2, pp. 48–60. ISSN: 0346-1238. DOI: 10.1080/03461238.1952.10414182.

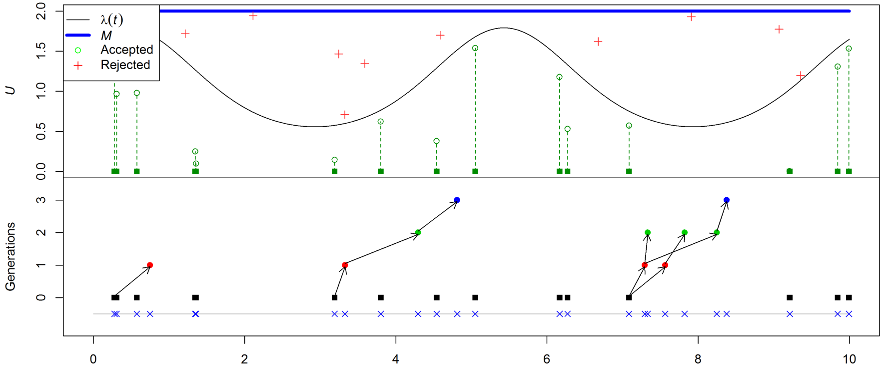
Simulate Hawkes in *R* (Ogata, 1981)

```
sim <- hawkes(T=10, fun=1, repr=1, family=''exp'', rate=2)
plot(sim, intensity = TRUE)
```

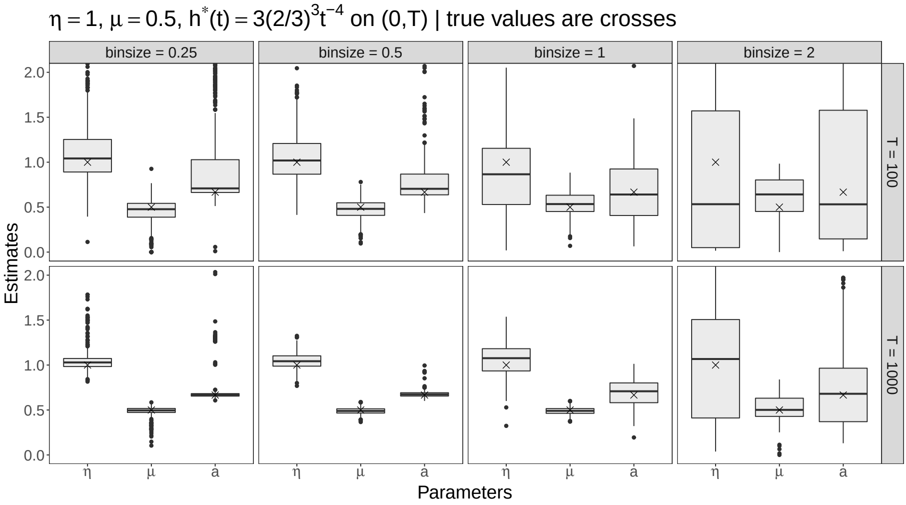


Simulate Hawkes with inhomogeneous background intensity in *R* (Møller and Rasmussen, 2005; Dassios and Zhao, 2013)

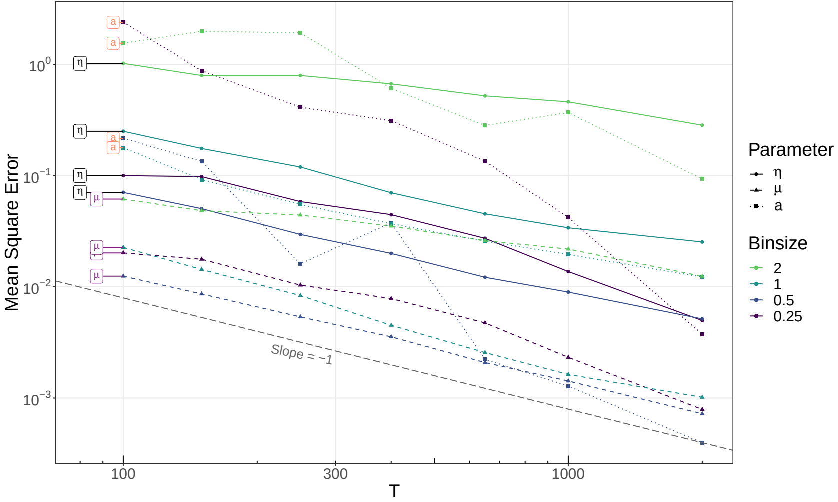
```
int <- function(t) exp(.5*cos(2*pi*t/5)+.3*sin(2*pi*t/5))
sim <- hawkes(T=10, fun=int, M=2, repr=1, family=''exp'', rate=
plot(sim$immigrants)
plot(sim)
```



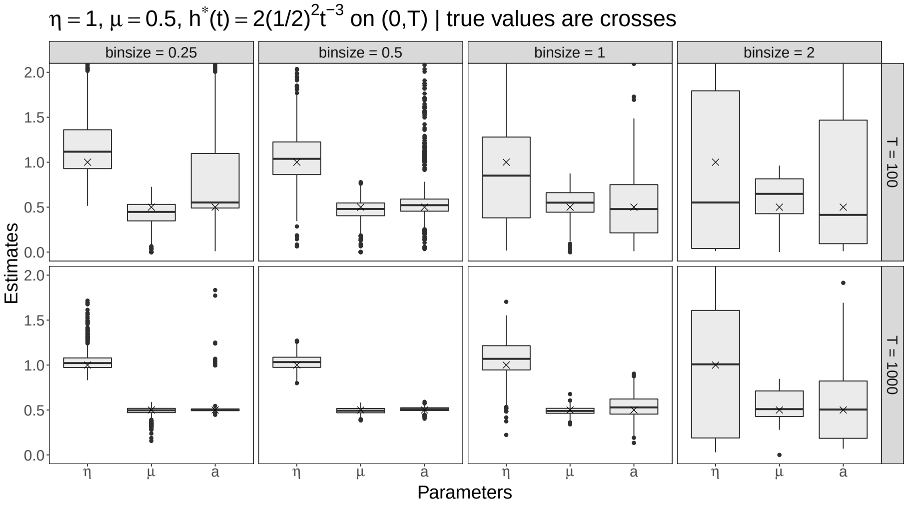
Pareto kernel, $\gamma = 3$



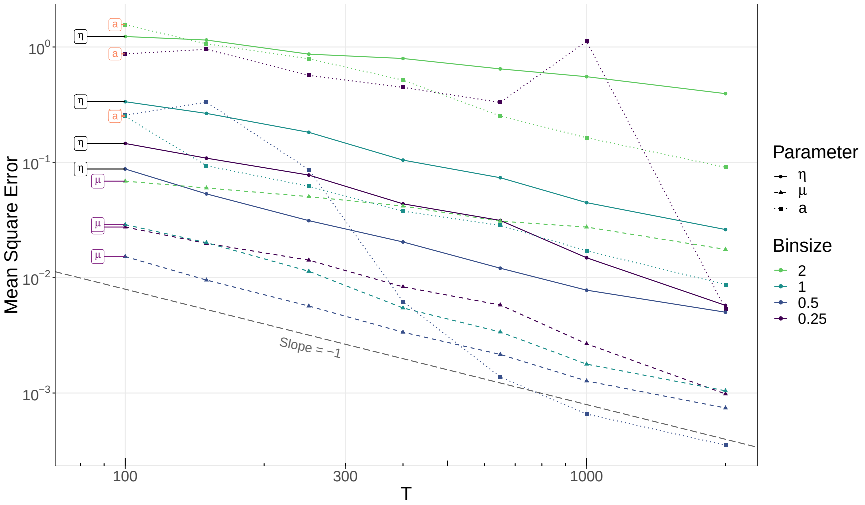
Pareto kernel, $\gamma = 3$



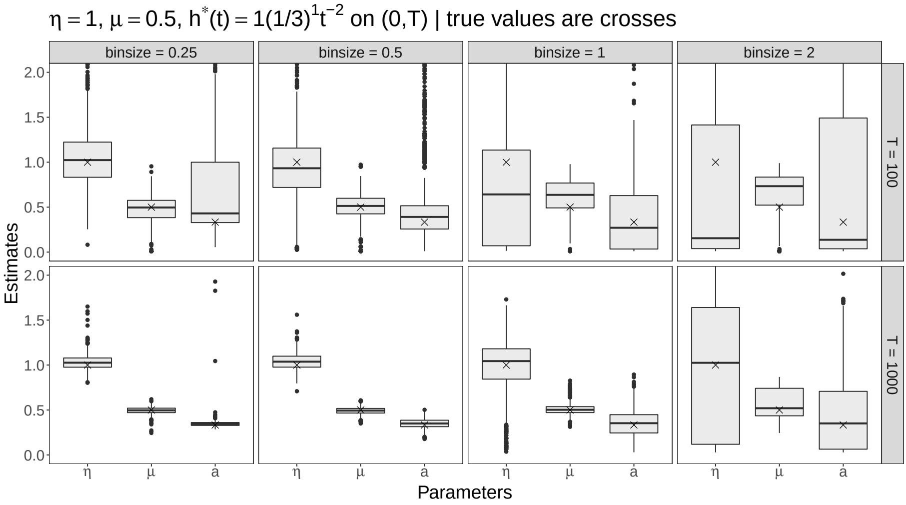
Pareto kernel, $\gamma = 2$



Pareto kernel, $\gamma = 2$



Pareto kernel, $\gamma = 1$



Pareto kernel, $\gamma = 1$