A strong mixing condition for Hawkes processes and its application to Whittle estimation from count data

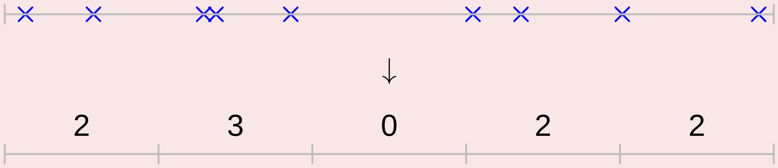
Felix Cheysson[1,2,3] Gabriel Lang[1]

[1] Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris, 75005, Paris, France.
[2] Epidemiology and Modeling of bacterial Evasion to Antibacterials Unit (EMEA), Institut Pasteur, Paris, France.
[3] Anti-infective Evasion and Pharmacoepidemiology Team, Centre for Epidemiology and Public health (CESP), Inserm / UVSQ, France.

Workshop: Statistical methods for Hawkes processes
March 10th, 2020

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.

Outline

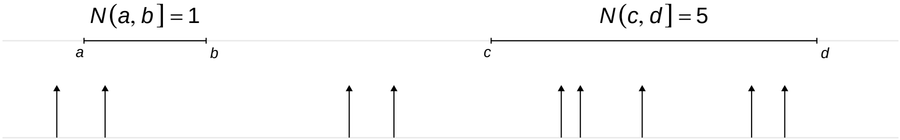
1. Spectral estimation of Hawkes processes
   - Notations
   - The Bartlett spectrum
   - Whittle estimation method
2. Strong mixing properties for Hawkes processes
   - Definitions
   - Strong mixing condition
   - Consequences for Whittle estimation
3. Simulation- and case-study
4. Conclusion and perspectives

Outline

Notations

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

Measurable map $N : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathfrak{N}, \mathcal{N})$ where $\mathfrak{N}$ denotes the set of locally finite counting measures on $\mathbb{R}$.



Consider a Hawkes process N on $\mathbb{R}$ (Hawkes, 1971) with intensity function:

$$\lambda(t) = \eta(t) + \int h(t-u) \mathrm{d}N(u)$$

where the reproduction function $h = \mu h^*$ has reproduction mean $\mu < 1$ and *kernel* h^* such that $\int h^* = 1$.

Spectral representation

Bartlett spectrum (Daley and Vere-Jones, 2008, Proposition 8.2.I)

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

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

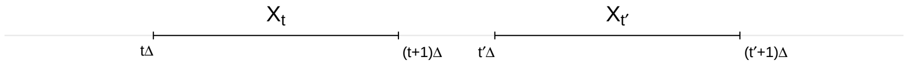
where φ and ψ are functions of rapid decay, $\psi^*(u) = \psi(-u)$, and $\widetilde{\cdot}$ denotes the Fourier transform: $\widetilde{\varphi}(\omega) = \int_{\mathbb{R}} e^{i\omega u}\varphi(u)du$.
The unique measure $\Gamma(\cdot)$ is called the *Bartlett spectrum* of N.

For the stationary Hawkes process N, the Bartlett spectrum admits a density given by (Daley and Vere-Jones, 2008, Example 8.2(e))

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

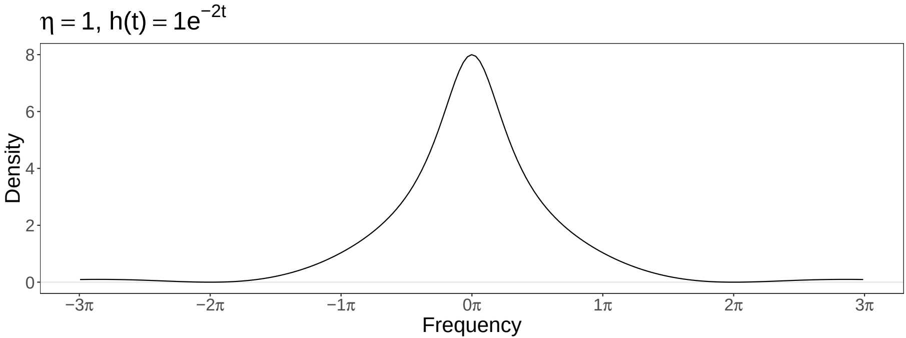
with $m = \mathbb{E}\left[N(0,1]\right] = \eta(1-\mu)^{-1}$.

Spectral representation of the bin-count process

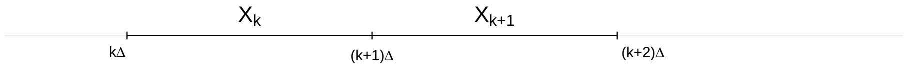


For the Hawkes bin-count process $\{X_t\}_{t \in \mathbb{R}} = \{N(t\Delta, (t+1)\Delta)\}_{t \in \mathbb{R}}$, the spectral density is given by

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

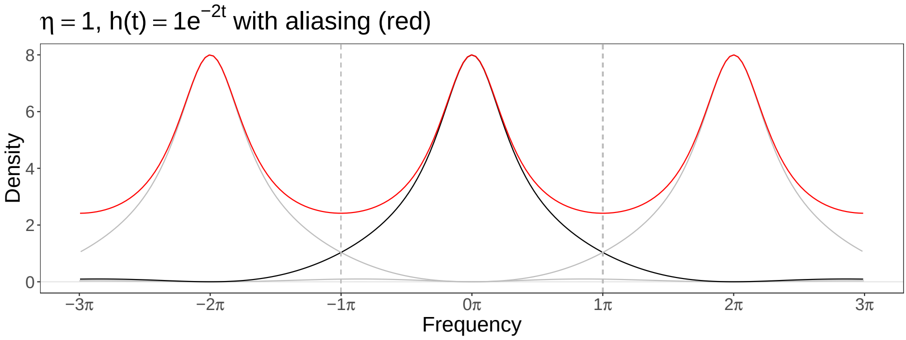


Spectral representation of the bin-count sequence



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

$$\widehat{\theta}_n = \operatorname*{arg\,min}_{\theta \in \Theta} \mathcal{L}_n(\theta)$$

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

- Started with (Rosenblatt, 1956), who defined the strong mixing coefficient to measure the dependence between random variables;

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

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

- Provides strong moment inequalities and coupling methods to achieve proof of asymptotic properties (Doukhan, 1994; Rio, 2017), provided the coefficient decreases fast enough.
- Difficult to bound in practice.
- 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 and random fields.

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

Dependence between past and future events:

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

Recent works: (Heinrich and Pawlas, 2013; Poinas, Delyon, and Lavancier, 2019).

Strong mixing properties for Hawkes processes

Theorem

Let N *be a stationary Hawkes process on* $\mathbb{R}$*. Suppose that there exists a* $\delta > 0$ *such that the distribution 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

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

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)

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

1. Control (1) by the covariance of counts.
   - Theorem 2.5 from (Poinas, Delyon, and Lavancier, 2019) using the association of Hawkes processes, as they are infinitely divisible (Evans, 1990, Theorem 1.1).
2. Rescale to a single branching process by conditioning on the cluster centre 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 cluster centre process.

Asymptotic properties of the Whittle estimator

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)

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:

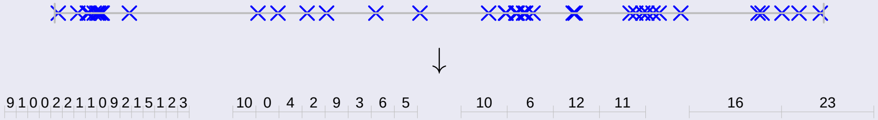
(A4) The function f_θ is twice differentiable with respect to θ and its second derivatives $(\partial^2/\partial\theta_k\partial\theta_l) f_\theta$ are continuous in $\theta \in \Theta$ and $-\pi \leq \omega \leq \pi$.

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

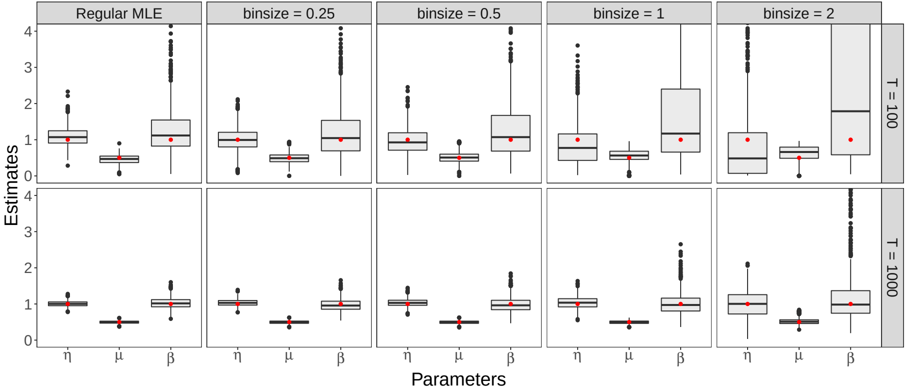
$$n^{1/2}(\widehat{\theta}_n - \theta_0) \underset{n\to\infty}{\sim} \mathcal{N}\left(0, \Gamma_{\theta_0}^{-1} + \Gamma_{\theta_0}^{-1} C_{4,\theta_0} \Gamma_{\theta_0}^{-1}\right).$$

Outline

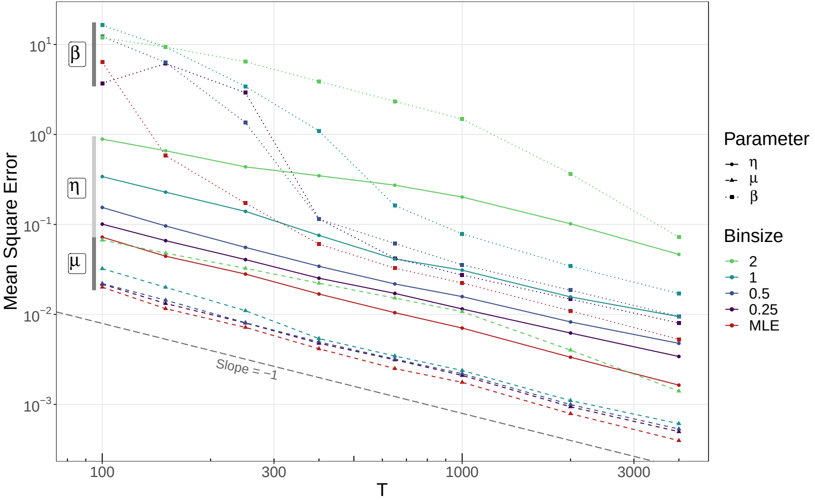
Simulation for the Whittle estimator



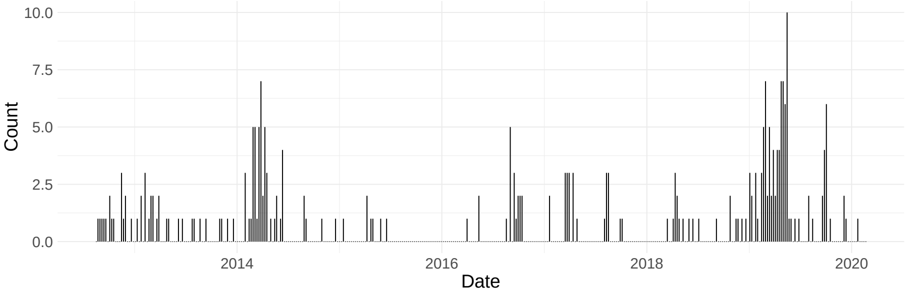
$\eta = 1,\ \mu = 0.5,\ h^{*}(t) = 1e^{-1t}$ on (0,T) | true values in red



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.

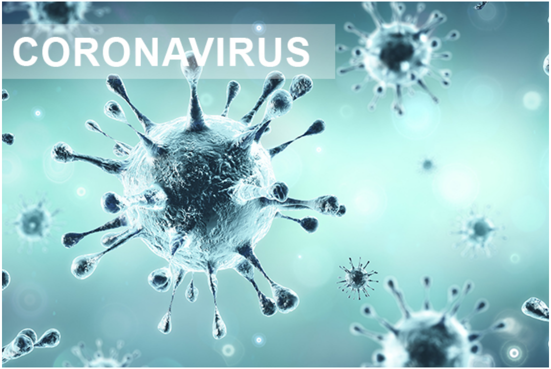
Simulation and estimation methods implemented in R package *hawkesbow*[2].

Direct extensions

- **Non causal Hawkes processes**: strong mixing properties do not care about causality.
- **Multivariate Hawkes processes**:
  - Multivariate Bartlett spectrum (Daley and Vere-Jones, 2003, Example 8.3(c));
  - Multitype Galton-Watson trees.

[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.

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: 00205958. DOI: 10.1007/BF01028982.

Bradley, Richard C. (2005). "Basic properties of strong mixing conditions. A survey and some open questions". In: *Probability Surveys* 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. 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.

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.

For Further Reading II

Daley, D. J. and David Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Probability and its Applications. New York: Springer. ISBN: 0-387-95541-0. DOI: 10.1007/b97277. arXiv: arXiv:1011.1669v3. URL: http://www.springerlink.com/content/978-0-387-21337-8http://link.springer.com/10.1007/b97277.

– (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. URL: http://projecteuclid.org/euclid.ecp/1465315601.

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

For Further Reading III

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. URL: http://epubs.siam.org/doi/10.1137/1119009.

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. URL: http://www.springerlink.com/index/D7X7KX6772HQ2135.pdfhttp://link.springer.com/10.1007/978-0-387-98135-2http://link.springer.com/10.1007/978-1-4612-4842-2.

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

For Further Reading IV

Hawkes, Alan G. (1971). "Point Spectra of Some Mutually Exciting Point Processes". In: *J. R. Stat. Soc. Ser. B* 33.3, pp. 438–443. ISSN: 00359246. DOI: 10.2307/2984686. arXiv: arXiv:1011.1669v3. URL: http://www.jstor.org/stable/2984686.

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. URL: http://dx.doi.org/10.1016/j.spa.2016.02.008http://linkinghub.elsevier.com/retrieve/pii/S0304414916000399.

For Further Reading V

- Møller, Jesper and Jakob G. Rasmussen (2005). "Perfect Simulation of Hawkes Processes". In: *Adv. Appl. Probab.* 37.3, pp. 629–646. URL: http://www.jstor.org/stable/30037347.
- 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. URL: http://ieeexplore.ieee.org/document/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. URL: http://arxiv.org/abs/1705.02276https://projecteuclid.org/euclid.bj/1560326425.

For Further Reading VI

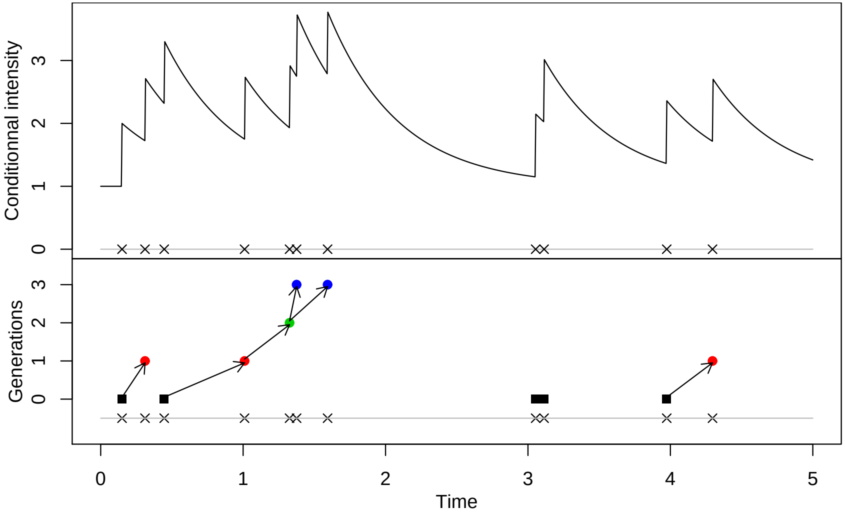
- Rio, Emmanuel (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Vol. 80. Probability Theory and Stochastic Modelling. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-662-54322-1. DOI: 10.1007/978-3-662-54323-8. URL: http://link.springer.com/10.1007/978-3-662-54323-8.
- Rosenblatt, 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. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.42.1.43.
- 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. URL: http://arxiv.org/abs/1704.01437https://projecteuclid.org/euclid.bj/1551862853.

For Further Reading VII

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. URL: http://projecteuclid.org/euclid.afm/1485893194.

Whittle, Peter (1952). "Some results in time series analysis". In: *Scand. Actuar. J.* 1952.1-2, pp. 48–60. ISSN: 16512030. DOI: 10.1080/03461238.1952.10414182.

Hawkes process as a branching process



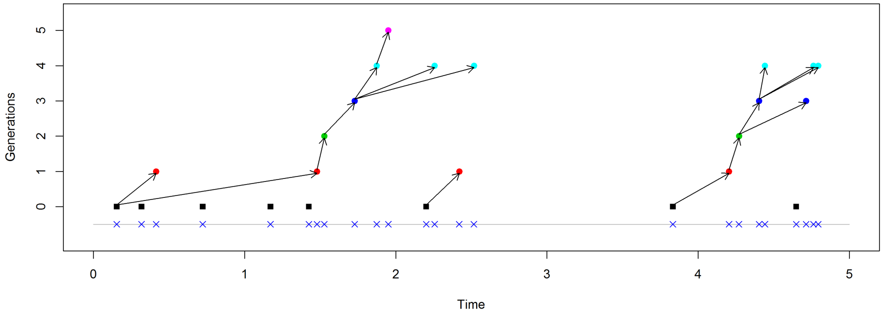
Epidemiological interpretation

Basic reproduction number

Mean number of infections caused by an individual

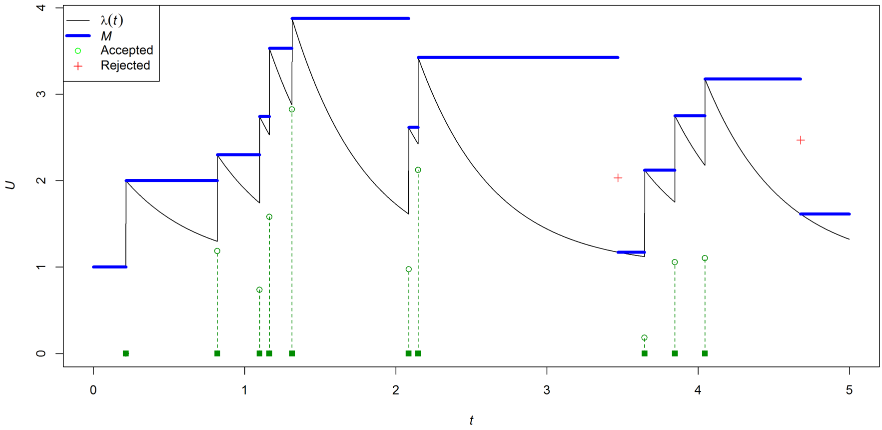
$$r = \int_0^\infty h(t)dt$$

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



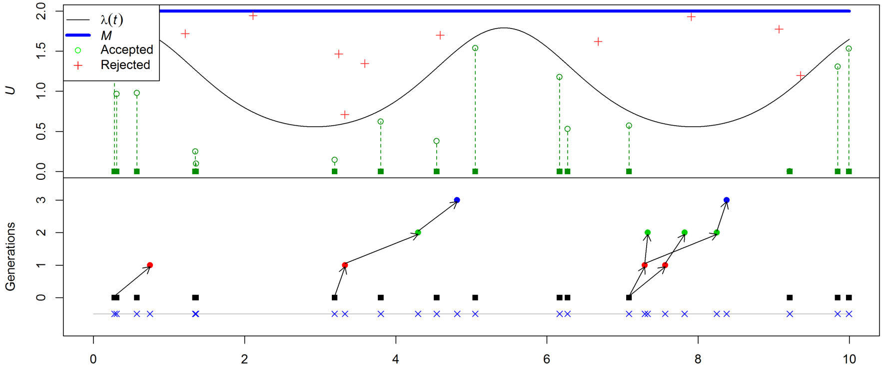
Simulate Hawkes in *R* (Ogata, 1981)

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

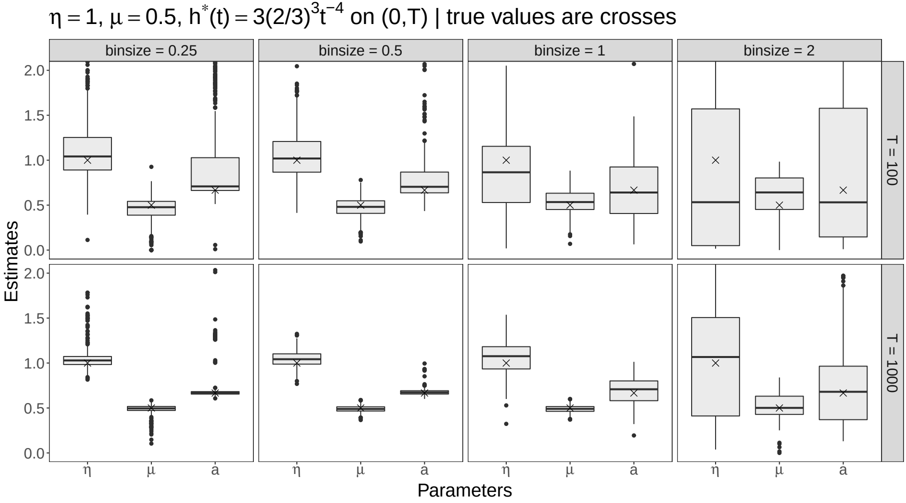


Simulate Hawkes with inhomogeneous background intensity in *R* (Møller and Rasmussen, 2005; Dassios and Zhao, 2013)

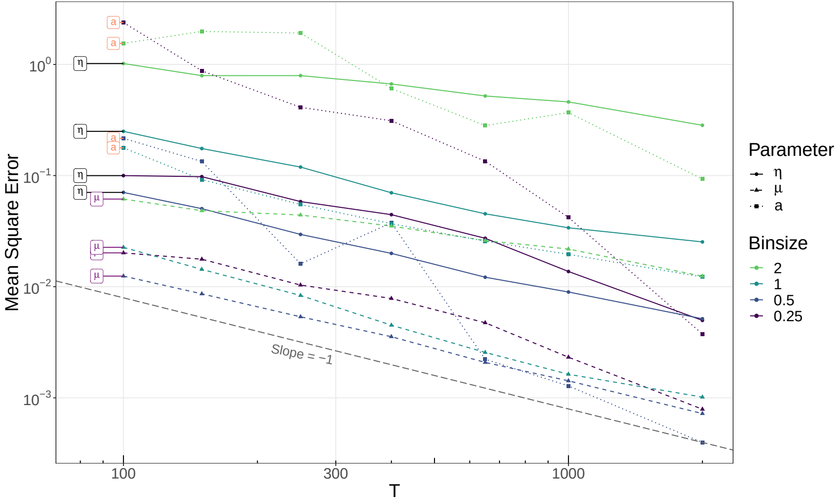
```
int <- function(t) exp(.5*cos(2*pi*t/5)+.3*sin(2*pi*t/5))
sim <- hawkes(T=10, fun=int, M=2, repr=1, family=''exp'', rate=
plot(sim$immigrants)
plot(sim)
```



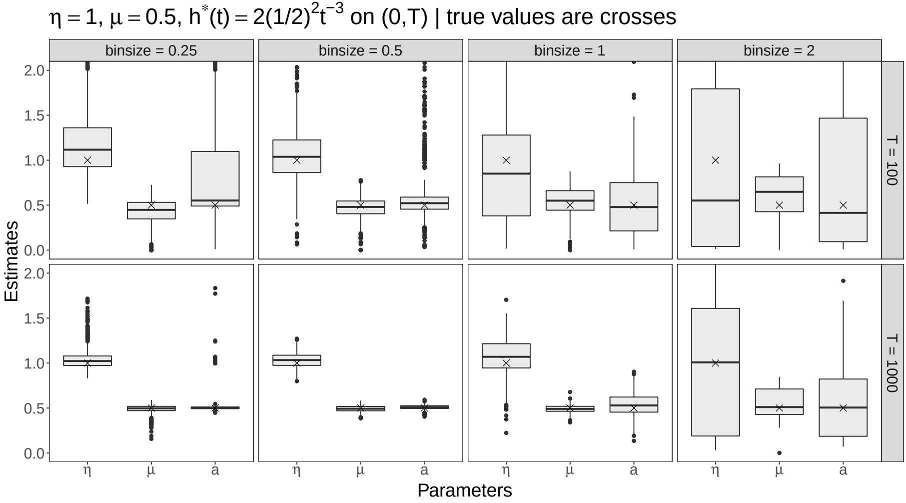
Pareto kernel, $\gamma = 3$



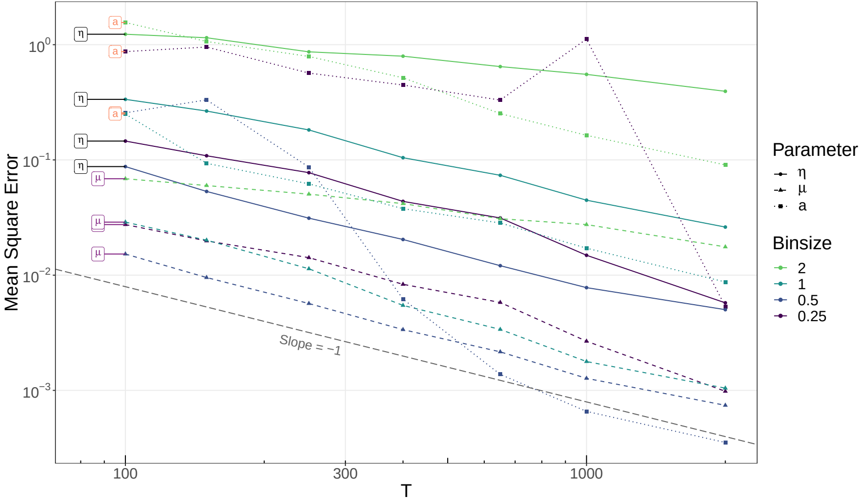
Pareto kernel, $\gamma = 3$



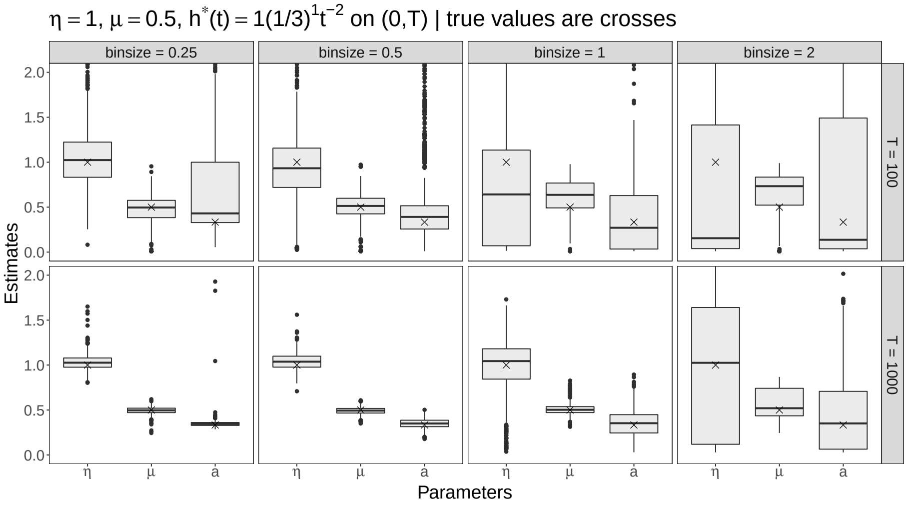
Pareto kernel, $\gamma = 2$



Pareto kernel, $\gamma = 2$



Pareto kernel, $\gamma = 1$



Pareto kernel, $\gamma = 1$