A causal framework for the estimation of attributable risks from aggregate data

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

[1] Sorbonne Université, LPSM
[2] AgroParisTech, MIA-Paris
[3] Institut Pasteur, Inserm, EMEA

Séminaire de statistique du MAP5
Vendredi 5 mars 2021

Risk factor and attributable risk

Risk factor : any variable associated with an increased risk of disease.

- Evidenced by epidemiological surveys (*e.g.* cohort, case-control studies) :

	Exposed	Unexposed
Cases	a	b
Controls	c	d

- Attributable risk (Levin, 1953) :

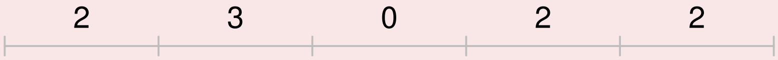
$$\mathrm{RA} = \frac{\mathbb{P}(D) - \mathbb{P}(D|\bar{E})}{\mathbb{P}(D)}.$$

- Can be interpreted as the proportion of cases that could be avoided if the exposure to the risk factor was removed.
- Estimator of the attributable risk (Benichou, 2001 ; Bard et al., 2005) :

$$\widehat{\mathrm{RA}} = \frac{ad - bc}{(a+b)(b+d)}.$$

Generalised linear models

Problem : aggregate data



Generalised linear model :
Let y_t denote the number of cases, x_t the level of exposure to the risk factor and Z_t some potential confounding factors ;

$$\begin{aligned} y_t &\sim \mathcal{L}_\theta(\mu_t), \\ g(\mu_t) &= \beta x_t + Z_t B. \end{aligned}$$

Attributable risk

Problem : the probabilities $\mathbb{P}(D)$ and $\mathbb{P}(D|\bar{E})$ are not well defined.
Idea : Define the attributable risk as the proportion of cases that would be prevented if the exposure to the risk factor was removed.

- Let :

$$y_t = y_t^* + \chi_t,$$

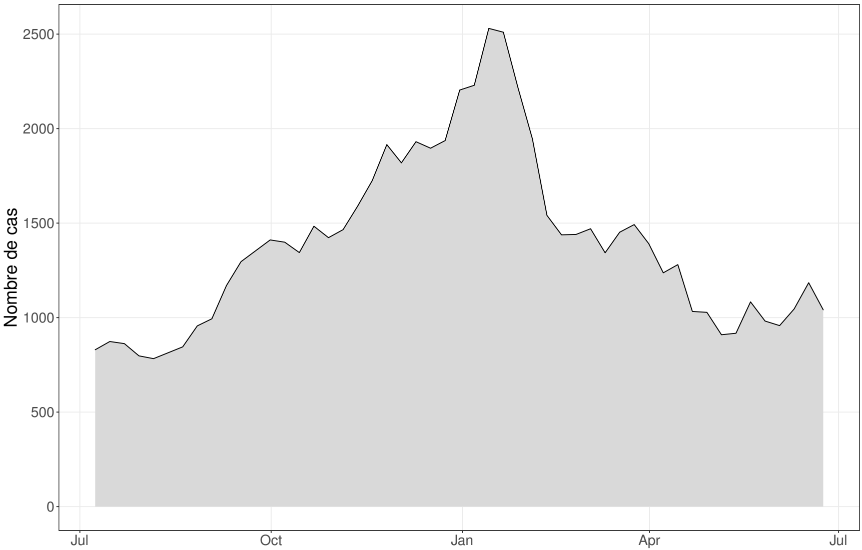
where y_t^* = number of cases in the absence of exposure ;,
and χ_t = contribution from the risk factor.

- Attributable fraction on a given window of time $\mathcal{T} = \{t_1, t_2, \ldots\}$:

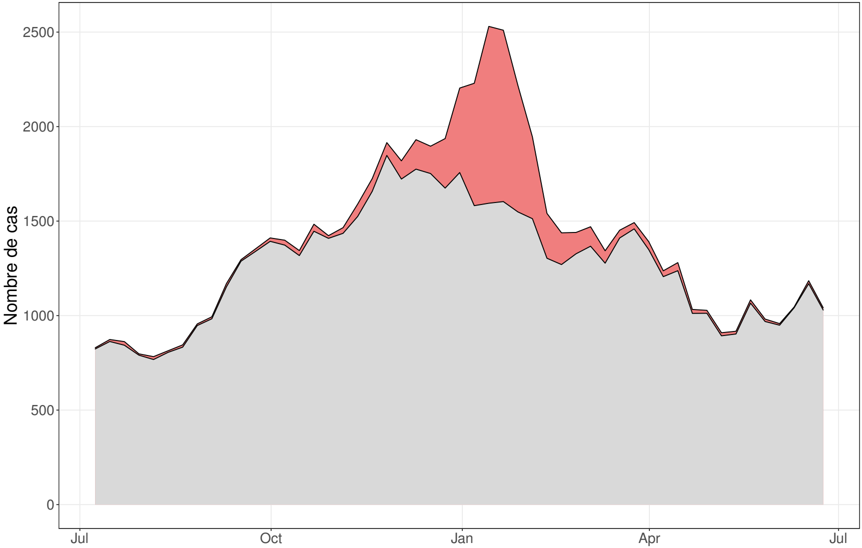
$$\mathrm{FA}(\mathcal{T}) = \frac{\sum_{t\in\mathcal{T}} y_t - y_t^*}{\sum_{t\in\mathcal{T}} y_t}.$$

- Objective :
  - Estimate the attributable risk conditionally to y_t and x_t ;
  - Equivalently, predict y_t^* (or χ_t) using the realisations of y_t and x_t.

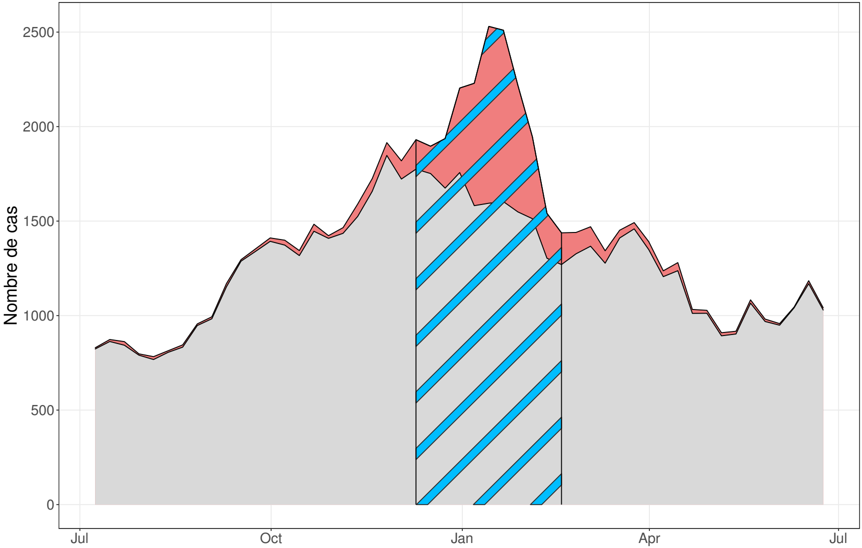
Attributable risk



Attributable risk



Attributable risk



Dependency between y_t and y_t^*

Estimating y_t^* by $\mathbb{E}[y_t^*]$ ignores the dependency between the baseline y_t^* and the outcome y_t.

Marginal distribution of y_t^* :

$$\begin{aligned} y_t^* &\sim \mathcal{L}_\theta(\mu_t^*), \\ g(\mu_t^*) &= Z_t B. \end{aligned}$$

For the linear regression model :

- Let :

$$y_t = \beta x_t + Z_t B + \varepsilon_t, \qquad y_t^* = Z_t B + \varepsilon_t.$$

- $y_t - y_t^* = \beta x_t,$ but $y_t - \mathbb{E}[y_t^*] = \beta x_t + \varepsilon_t.$

Some simple estimators

For the linear regression model :

$$\widehat{\mathrm{FA}}_n(\mathcal{T}) = \widehat{\beta}_n \frac{\sum_{t\in\mathcal{T}} x_t}{\sum_{t\in\mathcal{T}} y_t}.$$

For the Poisson regression model :

$$y_t \sim \mathcal{P}(\mu_t),$$
$$g(\mu_t) = \beta x_t + Z_t B.$$

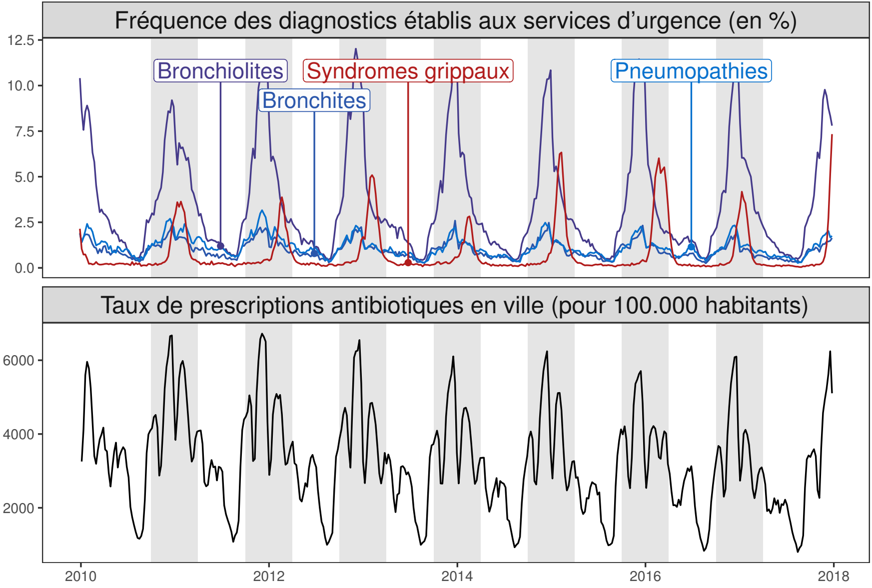
- Non additive error structure.
- Assume that $y_t^* \perp\!\!\!\perp \chi_t$.
- Then

$$\chi_t | y_t, x_t \sim B(y_t, p_t), \quad \text{où } p_t = \frac{\mu_t - \mu_t^*}{\mu_t}.$$

- And

$$\widehat{\mathrm{FA}}_n(\mathcal{T}) = \frac{\sum_{t\in\mathcal{T}} \widehat{p}_t y_t}{\sum_{t\in\mathcal{T}} y_t}.$$

Antibiotics and respiratory infectious diseases in children



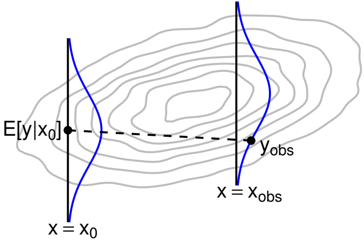
Estimated attributable fractions

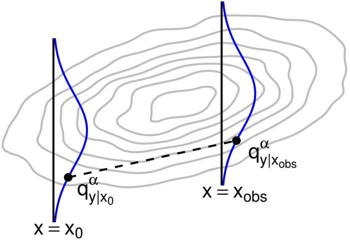
Fraction of antibiotic use attributable to lower respiratory infections : means [and 95% confidence intervals] during winter.

	75 years and above	5 years and below
Non viral lower respiratory infections		
Pneumonia	**38 [26–50]**	25 [19–32]
Viral lower respiratory infections		
Bronchiolitis		**19 [13–26]**
Bronchitis	17 [10–24]	14 [07–21]
Influenza-like illness	02 [01–03]	07 [05–09]

Observational *vs.* interventional distribution

- **Observational** $p(y \mid x)$: the distribution of Y given that X takes value x.
- **Interventional** $p(y \mid do(x))$: the distribution of Y if I were to set X at value x (Pearl, 2010 ; Tucci, 2013).





Functional causal models

Idea : dissociate the randomness on y from its dependency on x.

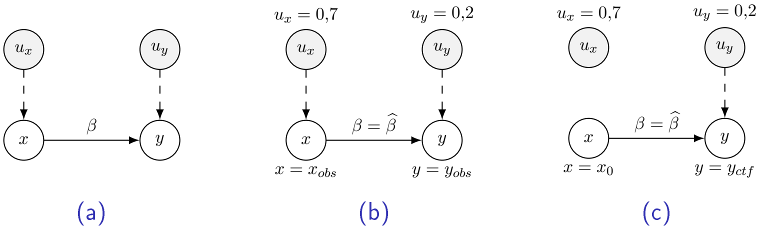


Figure – Causal model to answer the counterfactual question about a realisation $u = (u_x, u_y)$.

(a) The general model,

(b) The model specific to the observation $u = (0{,}7; 0{,}2)$,

(c) The mutilated model for which the exposure is set to $x = x_0$ according to the counterfactual question.

A counterfactual estimator

- Functional causal model :

$$x = u_x,$$
$$y = f_\beta(x, u_y),$$

- Attributable risk :

$$\mathrm{FA} = \frac{f_\beta(x, u_y) - f_\beta(x_0, u_y)}{f_\beta(x, u_y)}.$$

- Counterfactual estimator, *i.e.* specific to the observation :

$$\widehat{\mathrm{FA}} = \frac{y_{obs} - f_{\widehat{\beta}}(x_0, \widehat{u}_y)}{y_{obs}}.$$

- **Remark** : if $u_y \sim \mathcal{U}(0,1)$ and $f_\beta(x, u_y) = F^{-1}_{y|x}(u_y) = q^{y|x}_{u_y}$, then the numerator is the difference between the observed quantile of $p(y \mid x)$ and the estimated quantile of $p(y \mid x_0)$.

Motivation

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

Why the Hawkes process ?

- Usual autoregressive models can be difficult to interpret in a public health context.
- Low incidence diseases need discrete modelisation.

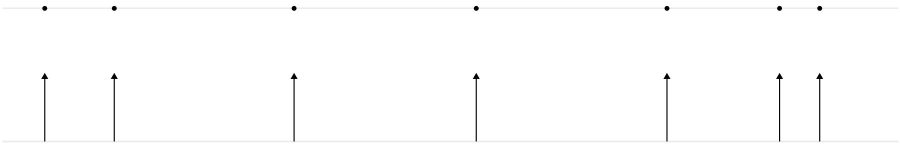
$\rightarrow$ Hawkes processes (Meyer, Elias et Höhle, 2012).

Point process

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

Point process

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



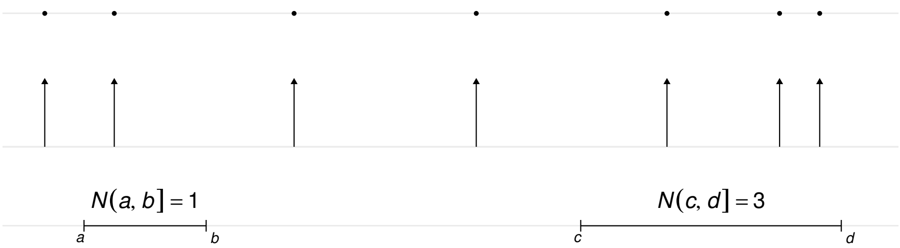
Point process

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 $\mathbb{R}$.



The 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 :

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

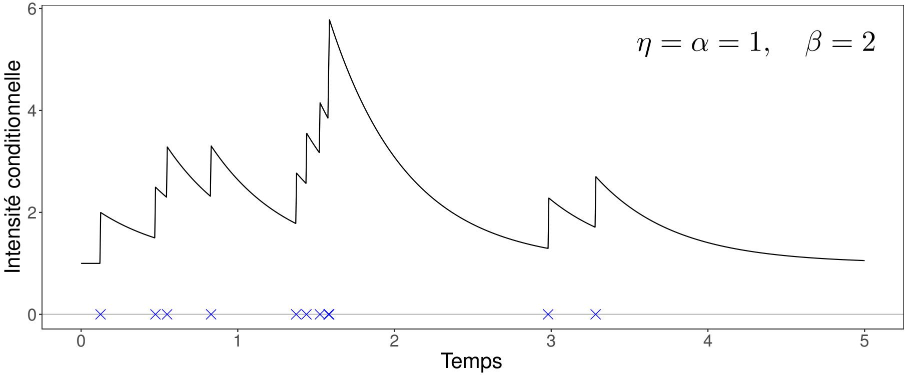
where η, h are integrable nonnegative functions such that $\int h < 1$, and $(t_j)_{j \in \mathbb{N}}$ are realisations of the point process.

The Hawkes process

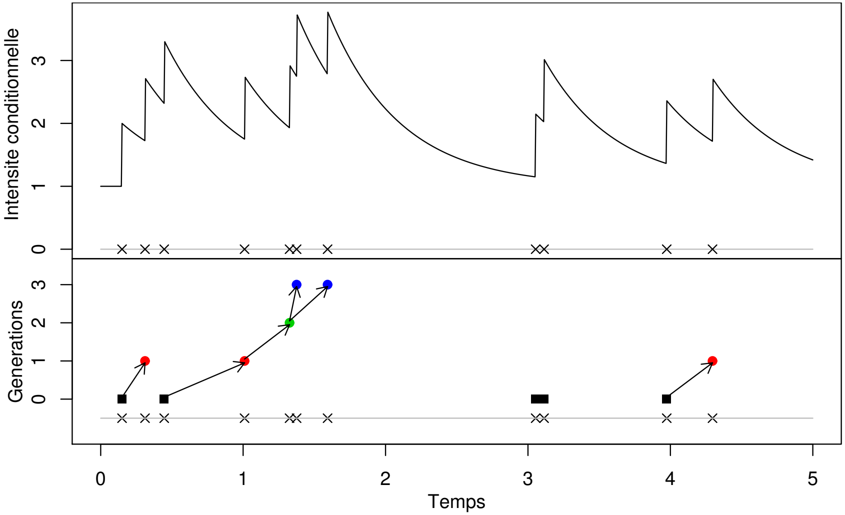
Linear Hawkes process on the real half-line

With exponentially decaying intensity :

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



The Hawkes process as branching processes



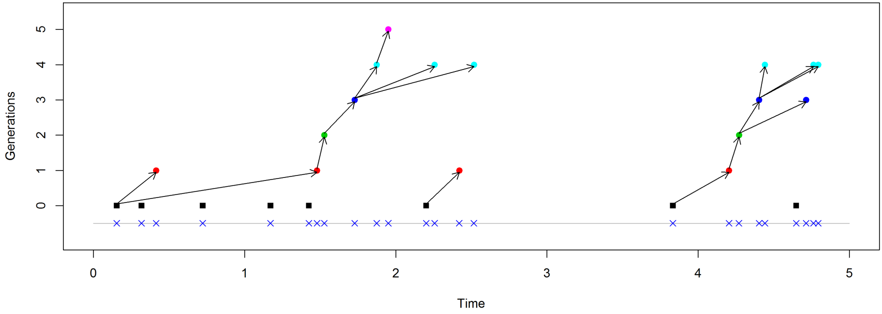
Epidemiological interpretation

Basic reproduction number

Mean number of infections caused by an individual

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

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



Attributable risk for the Hawkes process

- **Question** : How does a change in η or h impact the number of events ?
- **Partial answer** : For the mean, $\mathbb{E}[N(0,1)] = \eta/(1 - \int h)$.
  For the variance, (Da Fonseca et Zaatour, 2014 ; Daley et Vere-Jones, 2003).
- **Counterfactual estimator :**
  - **Idea** : Determine for each event the probability of it occurring under the new set of parameters.
  - **Problem** : Can't plug in new parameters into the conditional intensity ; some events depend on previous events occurring.

Probability that a point T_j is generated by T_i

$$\mathbb{P}(T_i \text{ parent of } T_j \mid T_i, T_j) = \frac{h(T_j - T_i)}{\lambda(T_j)}.$$

Poisson process embedding

Random time change theorem

If $(t_i)_{i\in\mathbb{N}}$ is a point process with conditional intensity $\lambda^*(t_i)$, and $s_i = \int_0^{t_i} \lambda^*(s)ds$, then $(s_i)_{i\in\mathbb{N}}$ is a unit rate Poisson process.

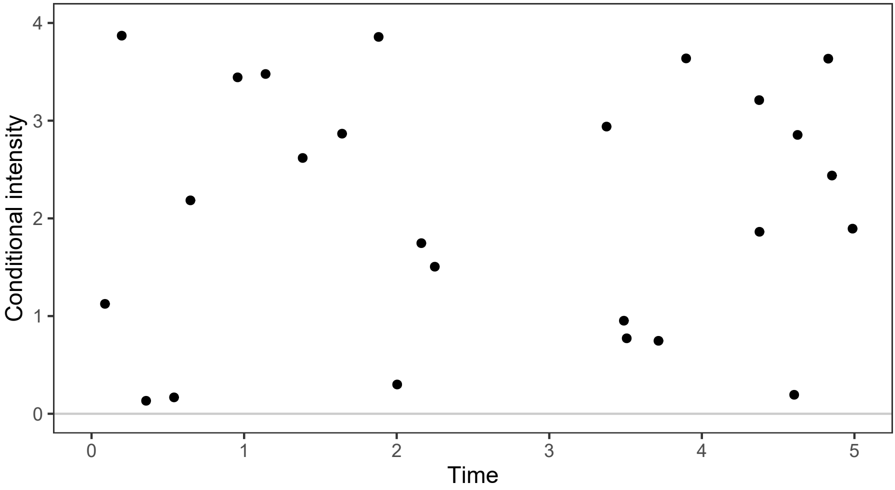
- Simulation algorithms via thinning Poisson processes (Lewis et Shedler, 1979 ; Ogata, 1981).
- Alternative construction via embedded Poisson process (Brémaud et Massoulié, 1996 ; Costa et al., 2018).

Let P denote a Poisson process with unit intensity on $\mathbb{R} \times (0, \infty)$. Then

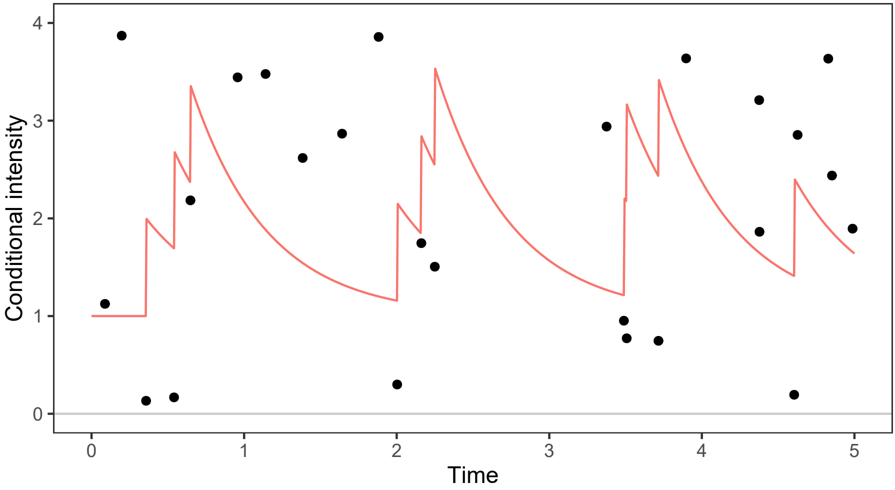
$$\begin{cases} N = \displaystyle\int_{\mathbb{R}\times(0,\infty)} \delta_u 1_{\{v\leq\lambda(u)\}} P(\mathrm{d}u, \mathrm{d}v), \\ \lambda(t) = \eta + \displaystyle\int_{\mathbb{R}} h(t-u) N(\mathrm{d}u), \end{cases}$$

is a Hawkes process with conditional intensity λ.

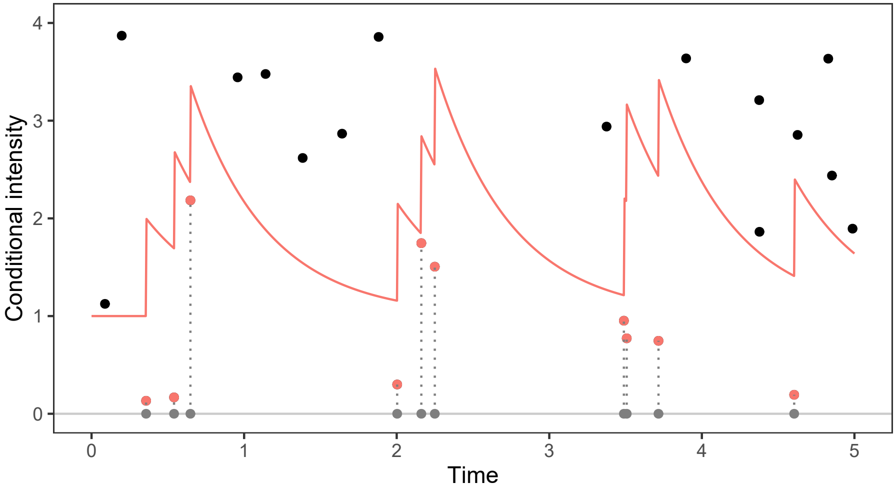
Poisson embedding for the Hawkes process



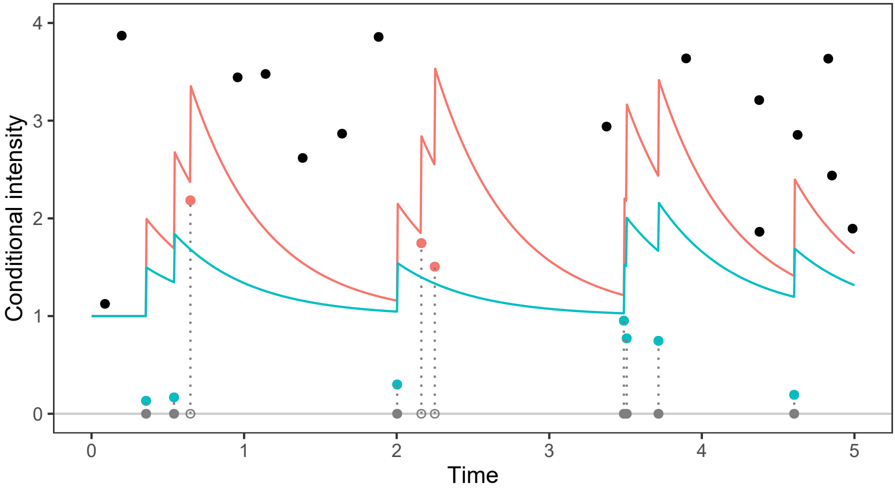
Poisson embedding for the Hawkes process



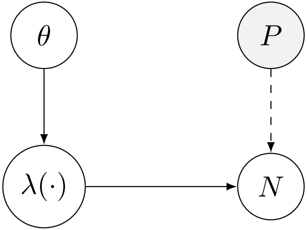
Poisson embedding for the Hawkes process



Poisson embedding for the Hawkes process



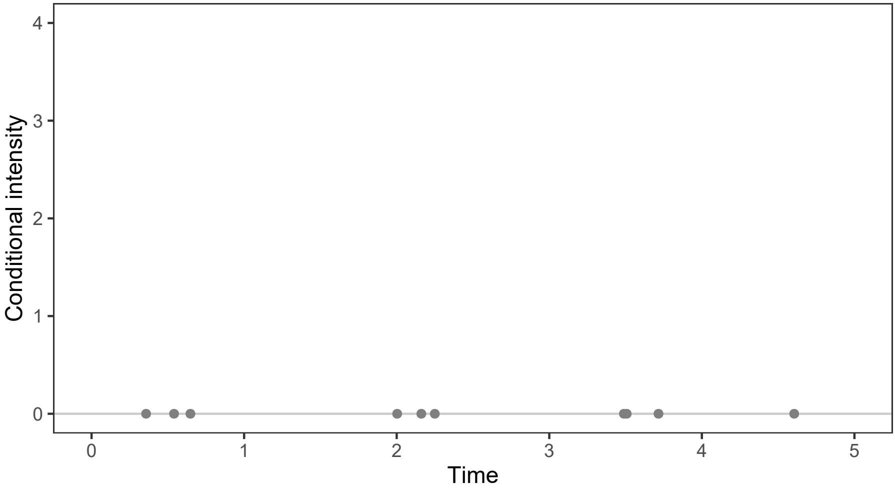
As a causal inference model



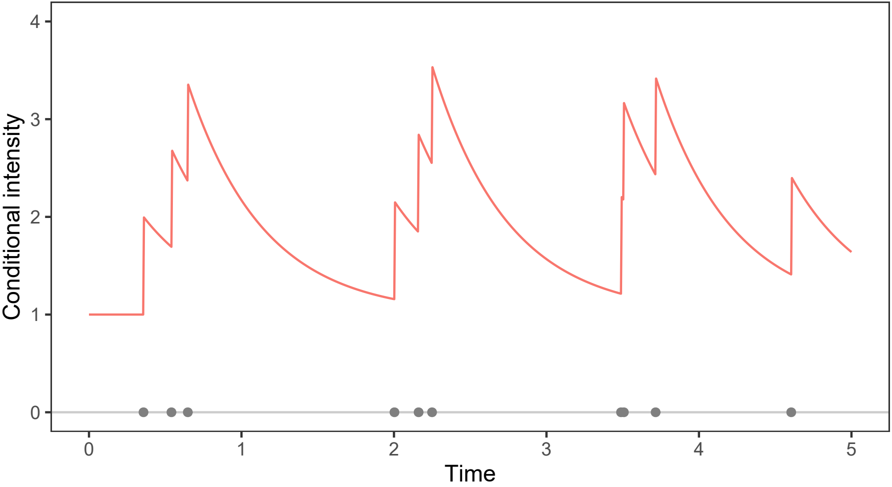
▷ Dissociate the random part P from the contribution of $\theta = (\eta, h)$:

$$\begin{cases} N = \displaystyle\int_{\mathbb{R}\times(0,\infty)} \delta_u 1_{\{v \leq \lambda_{\eta,h}(u)\}} P(\mathrm{d}u, \mathrm{d}v), \\ \lambda_{\eta,h}(t) = \eta + \displaystyle\int_{\mathbb{R}} h(t-u) N(\mathrm{d}u), \end{cases}$$

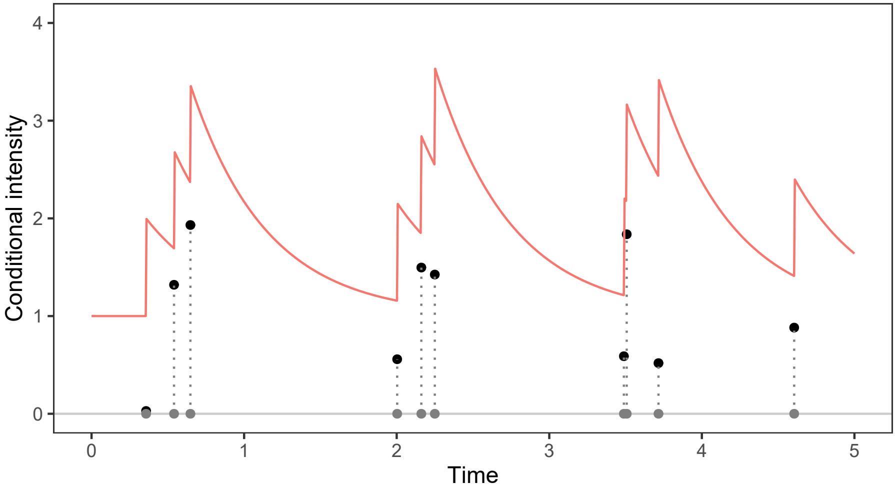
Monte Carlo estimation of attributable risk



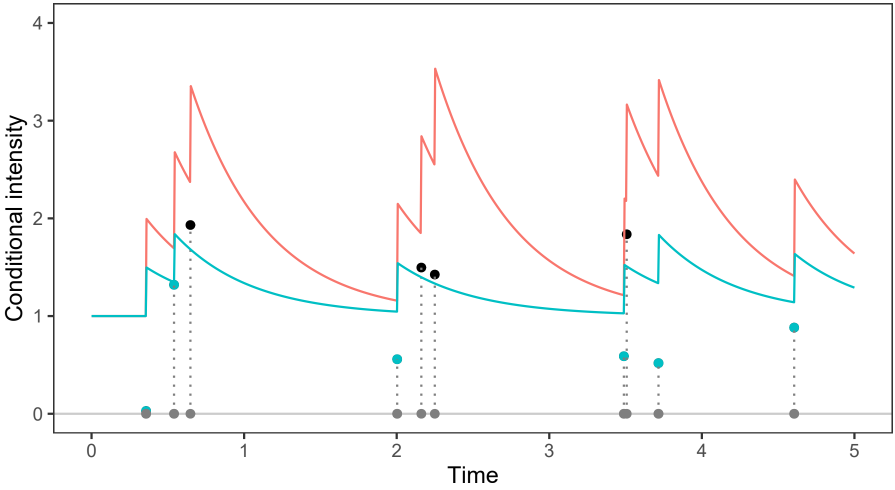
Monte Carlo estimation of attributable risk



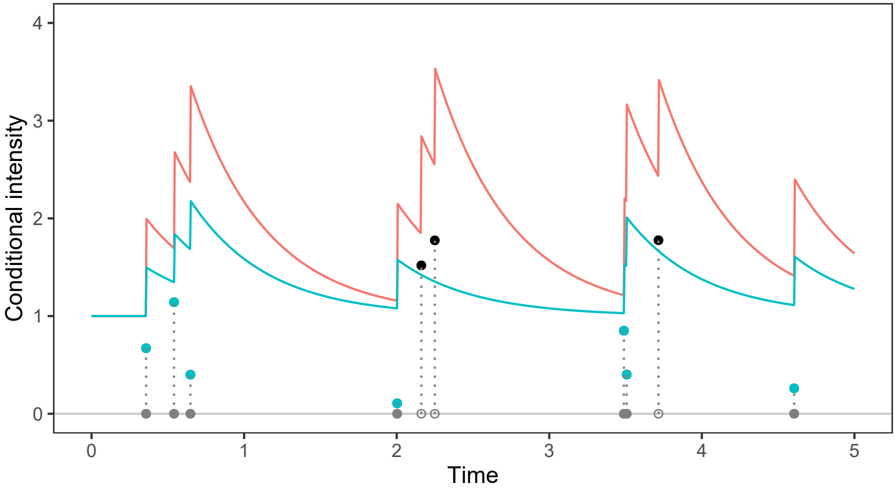
Monte Carlo estimation of attributable risk



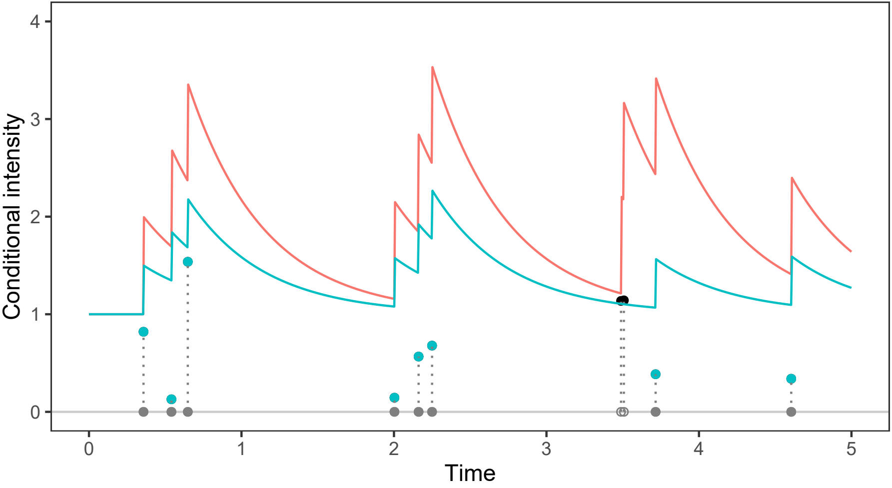
Monte Carlo estimation of attributable risk



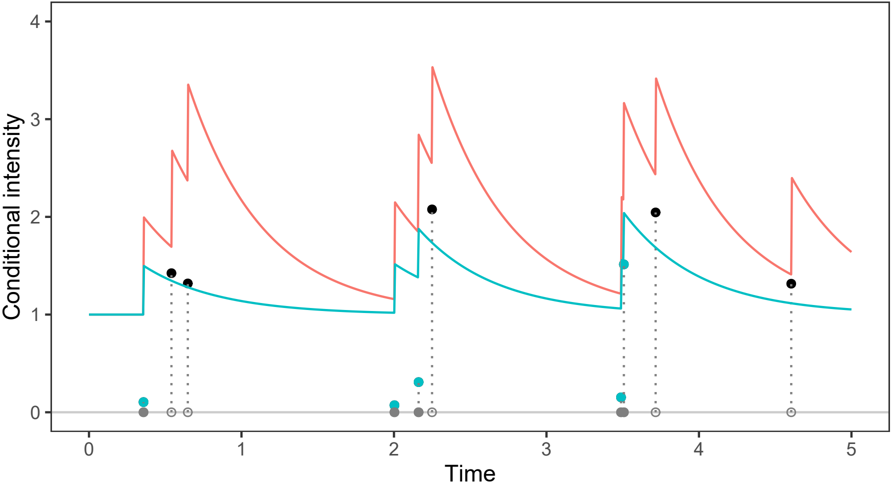
Multiple runs



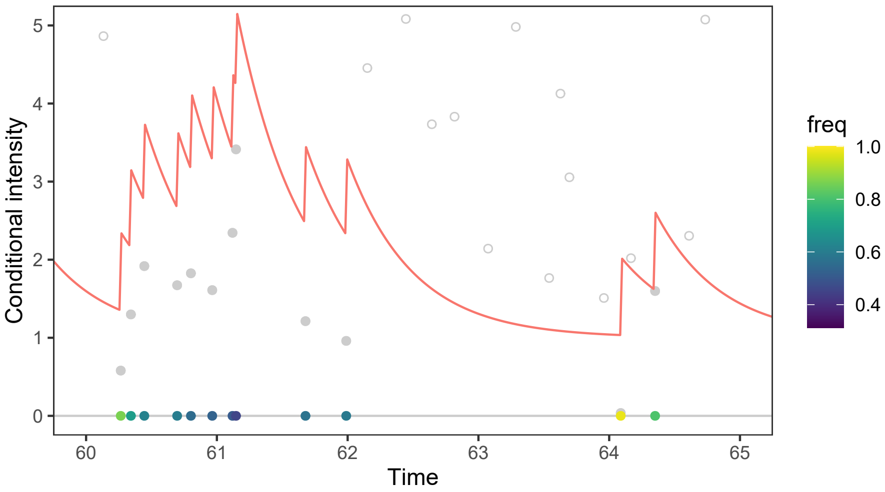
Multiple runs



Multiple runs



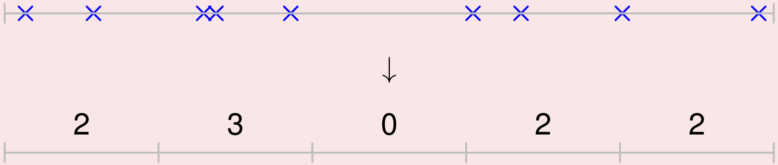
Attributable risk for each event



Perspectives

- Asymptotic properties ? Based on alpha-mixing properties of the Hawkes process (Cheysson et Lang, 2020).

Aggregate datasets



- Adequate estimation method (Kirchner, 2017 ; Cheysson et Lang, 2020).
- Estimate position of events (stochastic algorithms à la Metropolis-Hastings).

For Further Reading I

Bard, Denis et al. (2005). "Risque attribuable". In : *Cancer - Approch. méthodologique du lien avec l'environnement*. Les éditions Inserm, p. 69-92. isbn : 2-85598-844-6. url : http://www.ipubli.inserm.fr/bitstream/handle/10608/129/?sequence=1.

Benichou, Jacques (2001). "A review of adjusted estimators of attributable risk". In : *Stat. Methods Med. Res.* 10.3, p. 195-216. issn : 0962-2802. doi : 10.1177/096228020101000303.

Brémaud, Pierre et Laurent Massoulié (1996). "Stability of nonlinear Hawkes processes". In : *Ann. Probab.* 24.3, p. 1563-1588. issn : 0091-1798. doi : 10.1214/aop/1065725193. url : http://projecteuclid.org/euclid.aop/1065725193.

Cheysson, F. et G. Lang (2020). *Strong mixing condition for Hawkes processes and application to Whittle estimation from count data*.

For Further Reading II

Costa, Manon et al. (2018). "Renewal in Hawkes processes with self-excitation and inhibition". In : p. 1-36. arXiv : 1801.04645. url : http://arxiv.org/abs/1801.04645.

Da Fonseca, José et Riadh Zaatour (2014). *Hawkes process : Fast calibration, application to trade clustering, and diffusive limit*. T. 34. 6, p. 548-579. isbn : 6499219940. doi : 10.1002/fut.21644.

Daley, D. J. et 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. url : http://www.springerlink.com/content/978-0-387-21337-8http://link.springer.com/10.1007/b97277.

For Further Reading III

Dassios, Angelos et Hongbiao Zhao (2013). "Exact simulation of Hawkes process with exponentially decaying intensity". In : *Electron. Commun. Probab.* 18.62, p. 1-13. issn : 1083-589X. doi : 10.1214/ECP.v18-2717. url : http://projecteuclid.org/euclid.ecp/1465315601.

Hawkes, Alan G (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In : *Biometrika* 58.1, p. 83-90. issn : 00063444. doi : 10.2307/2334319. url : http://www.jstor.org/stable/2334319?origin=crossref.

Kirchner, Matthias (2017). "An estimation procedure for the Hawkes process". In : *Quant. Financ.* 17.4, p. 571-595. issn : 1469-7688. doi : 10.1080/14697688.2016.1211312. arXiv : 1509.02017. url : http://arxiv.org/abs/1509.02017https://www.tandfonline.com/doi/full/10.1080/14697688.2016.1211312.

For Further Reading IV

- Levin, Mark L (1953). "The occurrence of lung cancer in man". In : *Acta Unio int contra cancrum* 9, p. 531-941.
- Lewis, PA W et Gerald S Shedler (1979). "Simulation of nonhomogeneous Poisson processes by thinning". In : *Naval research logistics quarterly* 26.3, p. 403-413.
- Meyer, Sebastian, Johannes Elias et Michael Höhle (2012). "A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence". In : *Biometrics* 68.2, p. 607-616. issn : 0006341X. doi : 10.1111/j.1541-0420.2011.01684.x. arXiv : 1508.05740.
- Møller, Jesper et Jakob G. Rasmussen (2005). "Perfect Simulation of Hawkes Processes". In : *Adv. Appl. Probab.* 37.3, p. 629-646. url : http://www.jstor.org/stable/30037347.

For Further Reading V

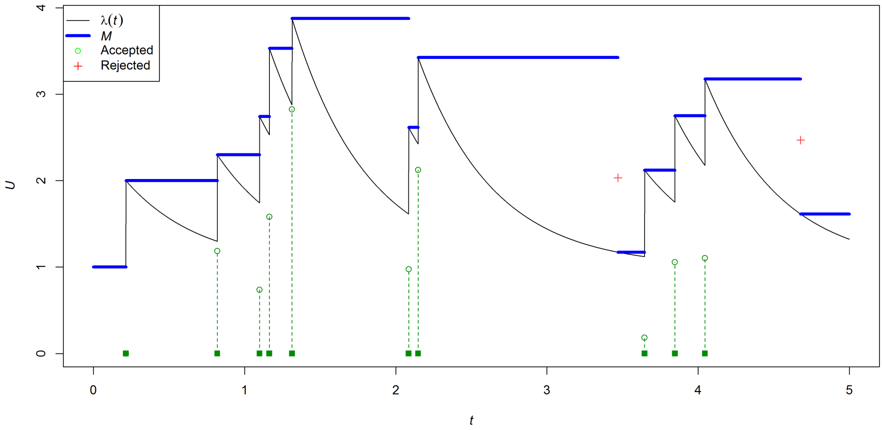
Ogata, Y. (1981). "On Lewis' simulation method for point processes". In : *IEEE Trans. Inf. Theory* 27.1, p. 23-31. issn : 0018-9448. doi : 10.1109/TIT.1981.1056305. url : http://ieeexplore.ieee.org/document/1056305/.

Pearl, Judea (2010). "The Foundations of Causal Inference". In : *Sociol. Methodol.* 40.1, p. 75-149. issn : 0081-1750. doi : 10.1111/j.1467-9531.2010.01228.x. url : http://journals.sagepub.com/doi/10.1111/j.1467-9531.2010.01228.x.

Tucci, Robert R. (2013). "Introduction to Judea Pearl's Do-Calculus". In : *arXiv e-prints*, arXiv :1305.5506. arXiv : 1305.5506. url : http://arxiv.org/abs/1305.5506.

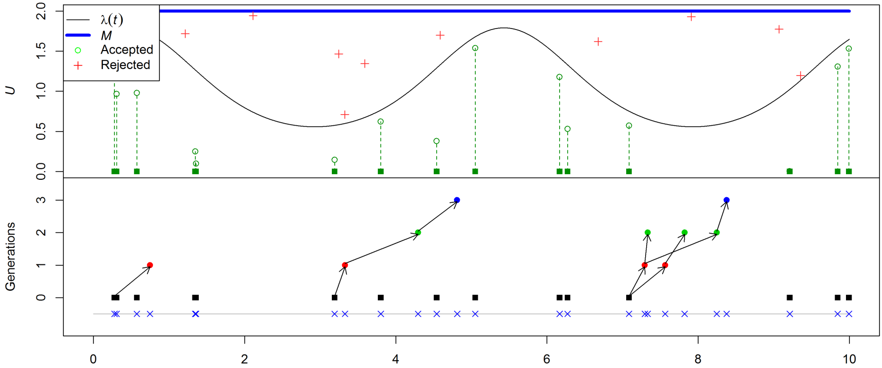
Simulate Hawkes in *R* (Ogata, 1981)

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

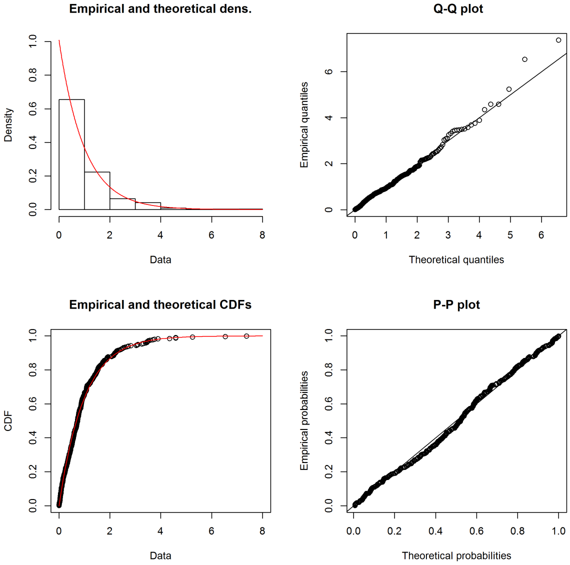


Simulate Hawkes with inhomogeneous background intensity in *R* (Møller et Rasmussen, 2005 ; Dassios et Zhao, 2013)

```
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)
```



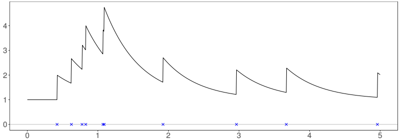
Residual analysis



Idea : spectral approach

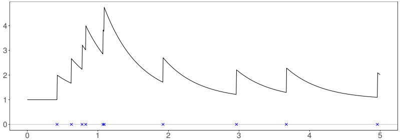
Objective : Estimate $\theta = (\eta, h)$ from the count process

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



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

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



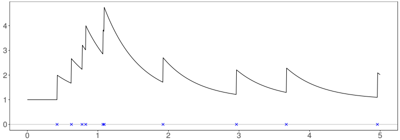
Likelihood of the count process is not tractable

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

Idea : spectral approach

Time domain

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



$\downarrow$

Likelihood of the count process is not tractable

$\downarrow$

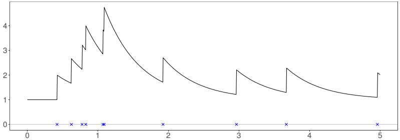
Objective : Estimate $\theta = (\eta, h)$ from the count process

Frequency domain

Idea : spectral approach

Time domain

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

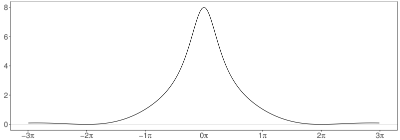


Likelihood of the count process is not tractable

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

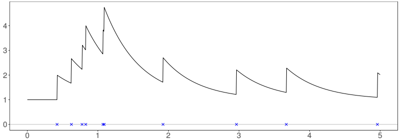
Frequency domain

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



Time domain

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

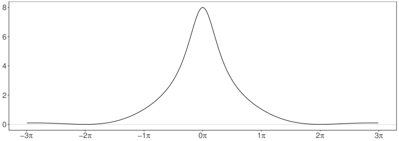


Likelihood of the count process is not tractable

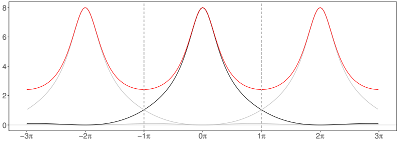
Objective : Estimate $\theta = (\eta, h)$ from the count process

Frequency domain

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



Spectral aliasing



Spectral density function : f_θ