Imperial College
London

25 September 2020

# Identifiability of Hawkes Processes

Andrew Connell

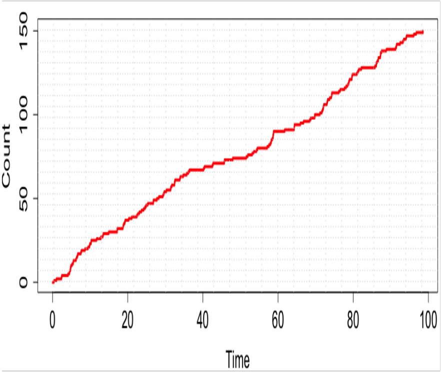
Supervisor: Dr. Ed Cohen
With Thanks to Leigh Shlomovich

# Introduction to Hawkes Processes

Imperial College
London

# Introduction to Hawkes Processes

- Hawkes processes are a special type of counting process. They model self-exciting behaviour [Hawkes, 1971].
- A Hawkes process can be one-dimensional or $M$-dimensional.
- There are three parameters to consider: $\mu, \alpha$ and $\beta$.

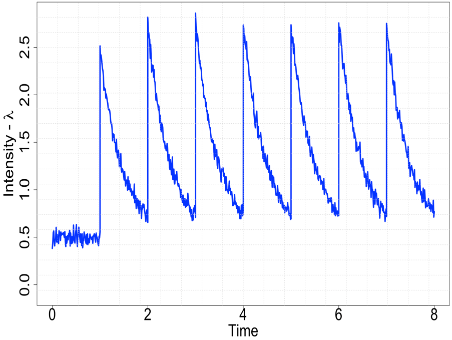


# The One-Dimensional Hawkes Processes

Imperial College London

# Intensity of Hawkes Processes

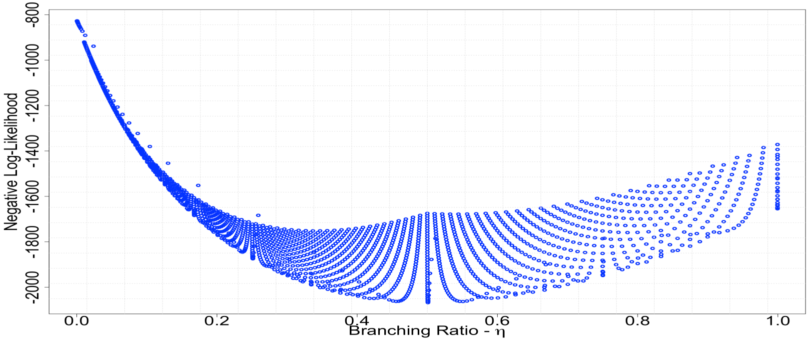
- Hawkes processes are completely identified by their intensity function.
- Given a $1^{st}$ order exponential kernel, the intensity of one-dimensional process is given below.




$$\lambda(t) = \mu + \int_0^t \alpha e^{-\beta(t-s)} \, \mathrm{d}N_i(s) = \mu + \sum_{t_i \leq t} \alpha e^{-\beta(t-t_i)}.$$

# Branching Ratio of a Hawkes Process

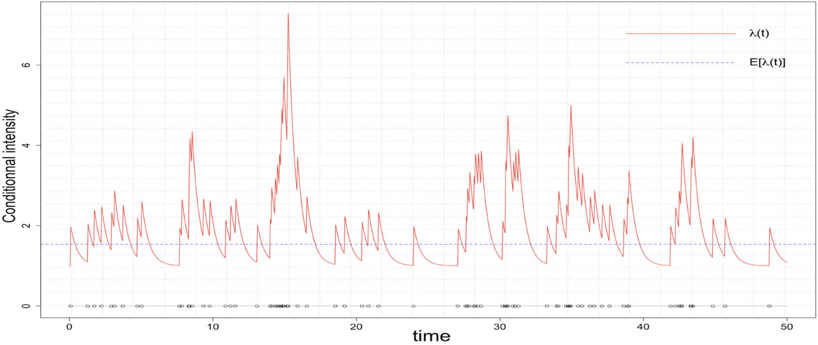
It is well known that the stationary condition of a Hawkes process is $\eta = \frac{\alpha}{\beta} < 1$ [Cordi et al., 2018].



# Expected Intensity

For the one-dimensional stationary Hawkes process the intensity is:

$$\mathbb{E}[\lambda(t)] = \frac{\mu}{1 - \alpha/\beta} = \frac{\mu}{1 - \eta}.$$



Imperial College London

# Likelihood of a Hawkes Process

- The log-likelihood can be defined recursively [Ogata, 1988].
- For the one-dimensional process with an exponential kernel: given the observations $t_1, ..., t_N$ on the interval $[0, t]$, the log-likelihood can be defined as

$$\ell(t|\theta) = \sum_{k=1}^{N} \log\left(\mu + \alpha R(k)\right) - \mu t - \frac{\alpha}{\beta} \sum_{k=1}^{N} \left(1 - e^{-\beta(t - t_k)}\right),$$

with $R(k) = e^{-\beta(t_k - t_{k-1})}(1 + R(k-1))$ and $R(1) = 0$.

# Introduction to Identifiability

# Identifiability

If a unique bijective mapping from the parameter space, $\Theta$, to the space of distributions for the data exists, then the model is identifiable [Patel et al., 2019]. Assume that some arbitrary data, $Y$, has log-likelihood $\ell(Y;\theta)$. A model is identifiable if for any $\theta, \phi \in \Theta$,

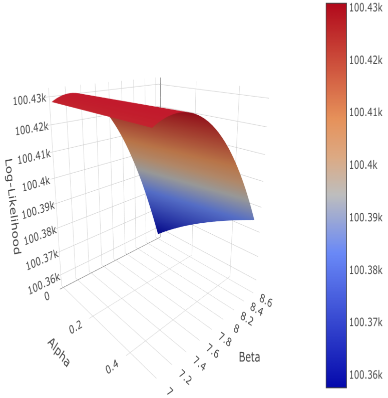
$$\ell(Y;\theta) = \ell(Y;\phi) \Rightarrow \theta = \phi.$$

Imperial College London

# Global Practical Identifiability

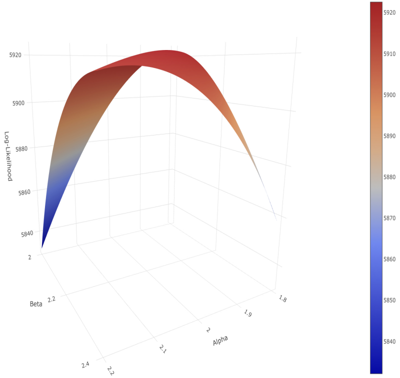
- Global identifiability means that the parameter set is unique on the parameter space being considered.

- Practical identifiability only considers the observations available.



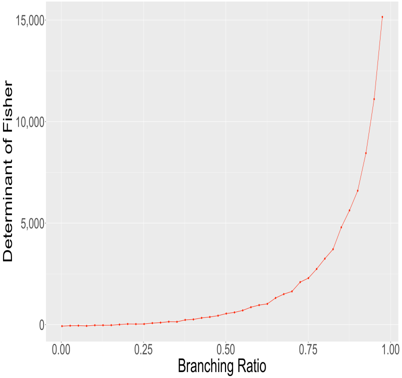
Imperial College London

# Identifiability of the Likelihood Surface




- Unidentifiability is often due to multiple parameters attaining the maximum likelihood.
- The maximum diagonal ridge occurs for parameters that share a branching ratio with the true parameters ($\gamma = \alpha/\beta = 2/2.2$).

# Identifiability and the Hessian




A process is locally identifiable if and only if the Hessian is non-singular [Rothenberg, 1971].

- The figure shows the use of MC estimation to find the determinant of the empirical Fisher Information.
- As the expected intensity increases the determinant also increases.

# The Importance of Identifiability

In much of the current literature identifiability is either assumed or ignored entirely. However,

- Unidentifiable processes do not have a unique solution.
- Unidentifiable processes lead to imprecise inference.

# Novel Methodology

# The Negative Moments

It is possible to extend Cui's method to find an iterative method for calculating the negative moments [Cui et al., 2020].

- For some $p < 0$, the negative moments of $N(t)$ are

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N^p(t)] = \sum_{j=p+1}^{0} \binom{-p}{-j} \mathbb{E}[N^j(t)\lambda(t)].$$

- For some $q < 0$, the negative moments of $\lambda(t)$ are

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^{q+1}(t)] = q\beta(\mathbb{E}[\lambda^{q+1}(t)] - \mu\mathbb{E}[\lambda^{q}(t)]) + \sum_{i=q+1}^{0} \binom{-q}{-i} \alpha^{i-q}\mathbb{E}[\lambda^{i}(t)].$$

# The Full Fisher Information

The one-dimensional Fisher information matrix utilises the Hessian [Ozaki, 1977] and can be denoted as

$$I(\theta) = \begin{pmatrix} \frac{(\beta-\alpha)t}{\beta\mu} & \frac{t}{\beta} & -\frac{\alpha t}{\beta^2} \\ \frac{t}{\beta} & \frac{\mu t}{\beta(\beta-\alpha)} + \frac{t}{2(\beta+\alpha)} & -\frac{\alpha\mu t}{\beta^2(\beta-\alpha)} \\ -\frac{\alpha t}{\beta^2} & -\frac{\alpha\mu t}{\beta^2(\beta-\alpha)} & \frac{\alpha^2\mu t}{\beta^3(\beta-\alpha)} \end{pmatrix}.$$

The closed-form for the $M$-dimensional case has also been derived and can be found in the thesis. The identifiability condition will follow from this result.

# The Determinant of the Fisher Information

The identifiability condition relies on the determinant of the Fisher information.

- For the one-dimensional Hawkes process,

$$\begin{aligned}\det(I(\theta)) = t^3\mathbb{E}[\lambda(t)]\bigg(&\frac{\alpha^2}{2\mu\beta^5}(\beta^2+2\mu-1)\\ &-\frac{\alpha^2}{2\mu\beta^4(\beta+\alpha)}(\beta^2-\mu(\alpha+\beta)-1)+\frac{\alpha^3}{\beta^6}\bigg).\end{aligned}$$

# Identifiability and the Fisher Information

The Fisher determinant must be non-negative. By rearranging the general formula, it is possible to find an identifiability condition.

**Theorem (Identifiability of Hawkes Processes)**

*For the one-dimensional Hawkes process,*

$$\mathbb{E}[\lambda(t)] \geq \frac{2\alpha + \beta(\beta^2 - \beta + 2)}{2(\alpha + \beta)(\beta^2 - 2\beta)},$$

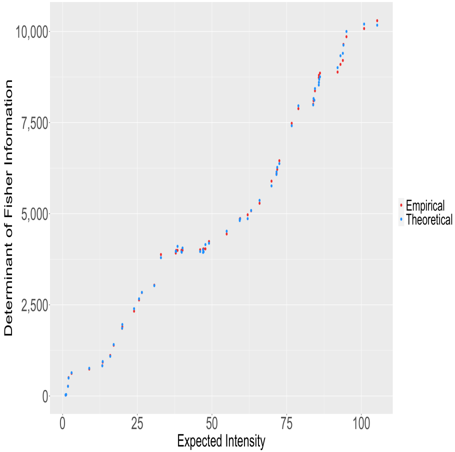
*provided the process is stationary. That is, $\alpha < \beta$.*

# Analysis

Imperial College London

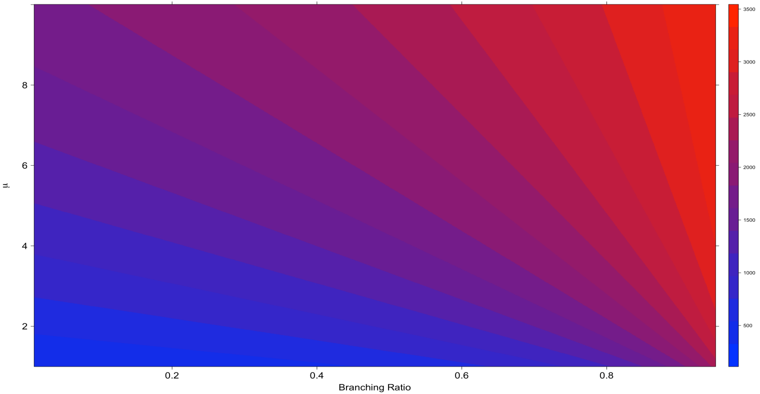
# Determinant of the Fisher information

- The Figure compares empirical estimates of the determinant through MC estimation with the theoretical results.
- Note, as the expectation increases so does the determinant.



# Parameter Effect on Identifiability

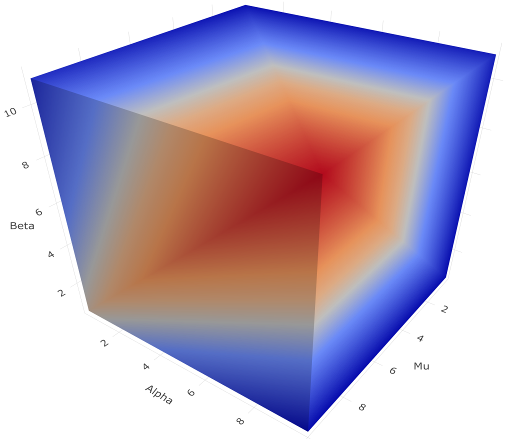
It is possible to form a heatmap of the determinant of the Fisher information by varying $\mu$ and the branching ratio, $\eta$.



Imperial College London

# Visualising Identifiability

- The identifiability condition can be visualised for the one-dimensional process.
- Red areas show parameter sets that are always identifiable.
- Blue areas show parameter sets that are always unidentifiable.
- White areas show parameter sets that sit on the boundary.



# The $M$-Dimensional Hawkes Processes

# The Full Hessian Matrix

The Hessian matrix of a Hawkes process can be generalised for an $M$-dimensional process with an exponential kernel of order $P$. In this setting, the parameter set, $\theta$, may be expressed as

$$\left( \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{pmatrix}, \left\{ \begin{pmatrix} \alpha_{1,j,1} & \alpha_{1,j,2} & \cdots & \alpha_{1,k,M} \\ \alpha_{2,j,1} & \alpha_{2,j,2} & \cdots & \alpha_{2,k,M} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{M,j,1} & \alpha_{M,j,2} & \cdots & \alpha_{M,k,M} \end{pmatrix} \right\}_{j=1}^{P}, \left\{ \begin{pmatrix} \beta_{1,j,1} & \beta_{1,j,2} & \cdots & \beta_{1,k,M} \\ \beta_{2,j,1} & \beta_{2,j,2} & \cdots & \beta_{2,k,M} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{M,j,1} & \beta_{M,j,2} & \cdots & \beta_{M,k,M} \end{pmatrix} \right\}_{j=1}^{P} \right).$$

The full thirty unique cases are given in the appendix of the thesis.

# Identifiability of the Fisher Information

For an $M$-dimensional Hawkes process,

$$\det(I(\theta)) = t^{M+2M^2}\left(\gamma \sum_{i=1}^{M} \mathbb{E}[\lambda_i(t)] + \delta\right).$$

Here, $\delta$ and $\gamma$ are terms to be found.

**Theorem (Identifiability of Hawkes Processes)**

*For the $M$-dimensional Hawkes process,*

$$\sum_{i=1}^{M} \mathbb{E}[\lambda_i(t)] \geq -\frac{\delta}{\gamma}.$$

# Concluding Remarks

# Conclusion

- All thirty elements of the $M$-dimensional Hessian now exists.
- By extending Cui's method, all moments can be iteratively calculated.
- The Fisher information for one-dimensional and $M$-dimensional Hawkes processes.
- The identifiability condition for Hawkes processes in the $M$-dimensional case may be expressed as

$$\sum_{i=1}^{M} \mathbb{E}[\lambda_i(t)] \geq -\frac{\delta}{\gamma}.$$

# References I

Cordi, M., Challet, D., and Toke, I. M. (2018).
Testing the causality of hawkes processes with time reversal.
*Journal of Statistical Mechanics: Theory and Experiment*, 2018(3):033408.

Cui, L., Hawkes, A., and Yi, H. (2020).
An Elementary Derivation of Moments of Hawkes Processes.
*Advances in Applied Probability*, 52(1):102–137.

Hawkes, A. G. (1971).
Spectra of some self-exciting and mutually exciting point processes.
*Biometrika*, 58(1):83–90.

Ogata, Y. (1988).
Statistical models for earthquake occurrences and residual analysis for point processes.
*Journal of the American Statistical Association*, 83(401):9–27.

Imperial College London

# References II


Ozaki, T. (1977).
Maximum likelihood estimation of hawkes' self-exciting point processes.
*Annals of the Institute of Statistical Mathematics*, 31(1):145–155.

Patel, L., Gustafsson, N., Lin, Y., Ober, R., Henriques, R., and Cohen, E. (2019).
A hidden markov model approach to characterizing the photo-switching behavior of fluorophores.
*Annals of Applied Statistics*, 13(3):1397–1429.

Rothenberg, T. J. (1971).
Identification in Parametric Models.
*Econometrica*, 39:577–591.