SIMULATION AND CALIBRATION OF A FULLY BAYESIAN MARKED MULTIDIMENSIONAL HAWKES PROCESS WITH DISSIMILAR DECAYS

Kar Wai Lim*,†, Young Lee†, Leif Hanlen†, Hongbiao Zhao‡

*Australian National University
†Data61/CSIRO
‡Xiamen University

1Proceedings of the 8th Asian Conference in Machine Learning (ACML) 2016
Introduction on Hawkes Processes

Simulation of Hawkes Processes

Bayesian Inference for Hawkes
Background

- **Poisson distributions**
  - Commonly used to model the number of times an event occurs in an interval of time or space.
  - Textbook example: the number of cars passing an intersection in half an hour.

![Poisson Distributions](image)

**Figure:** (left) Probability mass functions (right) Observed histogram

- **Additive Property:** If $X \sim \text{Poi}(\lambda_1)$, $Y \sim \text{Poi}(\lambda_2)$, then
  \[ X + Y \sim \text{Poi}(\lambda_1 + \lambda_2) \]
Background

- **(Homogeneous) Poisson process**
  - It is a stochastic process that keep track of the running counts of an event over time (and space).
  - For example, the number of cars passing an intersection is an evolution of counts with time:

![A Poisson process sample path](image)

- Call the evolution of counts as the counting process \( N(t) \) and the times of an event happening the event times \( t_i \).
Properties of Poisson process

- The counting process starts at zero: $N(t = 0) = 0$.

- Parameterised by the expected number of events per unit time, e.g. $\lambda = 3$ vehicles per minute.

- The counting process $N(t)$ at time $t$ follows $\text{Poi}(\lambda t)$. (number of events observed until time $t$)

- The difference (also called increment) in counting processes $N(t) - N(s) \sim \text{Poi}(\lambda(t - s))$ for $t > s$

Superposition property: If $N(t) \sim \text{PP}(\lambda_1)$, $M(t) \sim \text{PP}(\lambda_2)$, then $N(t) + M(t) \sim \text{PP}(\lambda_1 + \lambda_2)$
Background

What if some events are more frequent at certain times?

- More cars during peak hours!
- Instead of constant intensity, allow the intensity to vary with time: $\lambda(t)$ becomes a function of time.
Background

- **Extension: Inhomogeneous Poisson process**
  - **Example $\lambda(t)$:**
    - Piecewise linear;
    - Piecewise polynomial;
    - Cyclical functions such as sine curve.

**Figure:** Generated data
Background

- Properties of inhomogeneous Poisson process (IPP)
  - The counting process starts at zero: \( N(t = 0) = 0 \).
  - Parameterised by intensity function \( \lambda(t) \).
  - The counting process at time \( t \) follows \( \text{Poi} \left( \int_0^t \lambda(u) \, du \right) \).
  - The difference (also called increment) in counting processes
    \[
    N(t) - N(s) \sim \text{Poi} \left( \int_s^t \lambda(u) \, du \right) \quad t > s
    \]
  - Superposition property still holds: If \( N(t) \sim \text{IPP}(\lambda_1(t)) \), \( M(t) \sim \text{IPP}(\lambda_2(t)) \), then
    \[
    N(t) + M(t) \sim \text{IPP}(\lambda_1(t) + \lambda_2(t))
    \]
Hawkes Processes

- Hawkes process is a point process in which an occurrence of an event triggers future events (self-excitation).
- Our formulation of Hawkes (univariate):

\[ \lambda(t) = \mu(t) + \sum_{i=1: t > t_i}^{N(T)} Y_i e^{-\delta(t-t_i)} \]

- Decaying background intensity:

\[ \mu(t) = k + Y(0) e^{-\delta \times t} \]

- Random self excitations:

\[ Y_i \sim \text{i.i.d. Gamma} \]

- Terminology:
  - \( t_i, i = 1, \ldots, N(T) \) is a sequence of non-negative random variables such that \( t_i < t_{i+1} \), known as event times.
  - \( \Delta_i = t_i - t_{i-1} \) is called the inter-arrival time.
Multivariate Hawkes

- Captures multiple event types for which the events mutually excite one another.

- Our formulation (Bivariate Hawkes):

\[
\lambda_1(t) = \mu_1(t) + \sum_{j=1: t \geq t^1_j} Y_{1,j} e^{-\delta^1_1 t} + \sum_{j=1: t \geq t^2_j} Y_{1,j} e^{-\delta^1_1 t}
\]

\[
\lambda_2(t) = \mu_2(t) + \sum_{j=1: t \geq t^1_j} Y_{2,j} e^{-\delta^1_2 t} + \sum_{j=1: t \geq t^2_j} Y_{2,j} e^{-\delta^2_2 t}
\]

where \( \lambda_1(t) \) and \( \lambda_2(t) \) are the intensity functions for events 1 and 2, respectively.

- Note that the decay parameters \( \delta \) are different for each process.
Illustration of Multivariate Hawkes
Detour: Stationarity of Hawkes process

- Due to self-excitation property, a Hawkes process is only stable (stationary) when certain condition is satisfied.

- The intensity process $\lambda(t)$ explodes if this condition is not satisfied:
  - Causing chain reactions: intensity increases $\rightarrow$ more future events $\rightarrow$ further increases in intensity...

- We present a theoretical result on the expected stationary intensities for our Hawkes formulation. [Extension of Hawkes (1971) and Bacry et al. (2015)]
Outline

Introduction on Hawkes Processes

Simulation of Hawkes Processes

Bayesian Inference for Hawkes
Simulation of Hawkes Processes

- There are three categories of simulation methods.
- **1. Inverse Sampling (Ozaki, 1979)**
  - Derives cdf (cumulative distribution function) of inter-arrival times, then performs inverse sampling.
  - Cdf cannot be inverted directly so approximation is needed.
- **2. Thinning (Lewis and Shedler, 1979; Ogata, 1981)**
  - Simulate samples from a Poisson process and then *thin* the samples.
  - Akin to a rejection sampler.
- **3. Cluster method (Brix & Kendall, 2002; Møller & Rasmussen, 2005)**
  - Recast Hawkes using a Poisson cluster representation.
  - Each observed event generates an IPP.
  - Superposition of all of them forms a Hawkes process.
- **Notable mention: exact sampler of Dassios & Zhao (2013)**
  - Performs inverse sampling without approximation by decomposing a variable into two — need to satisfy a Markovian constraint.
- **Our method: exploits superposition theory and first order statistics for efficient sampling.**
Our Simulation Method in One Slide

- Illustration with bivariate Hawkes

\[
\lambda_1(t) = \mu_1(t) + \sum_{j=1: t \geq t_j^1} N_1(t) \cdot Y_{1,j} e^{-\delta_1^1 t} + \sum_{j=1: t \geq t_j^2} N_2(t) \cdot Y_{2,j} e^{-\delta_1^2 t}
\]

\[
\lambda_2(t) = \mu_2(t) + \sum_{j=1: t \geq t_j^1} N_1(t) \cdot Y_{2,j} e^{-\delta_2^1 t} + \sum_{j=1: t \geq t_j^2} N_2(t) \cdot Y_{2,j} e^{-\delta_2^2 t}
\]

- A Hawkes with intensity \( \lambda_1(t) \) is a superposition of IPP (with intensities \( \mu_1 \) etc).

- Inter-arrival times \((a_i, b_i, c_i...)\) for these IPP can be sampled easily.

- We show that the inter-arrival time \( \Delta_i \) for a Hawkes process is a first order statistics of these inter-arrival times:

\[
\Delta_i = \min\{a_i, b_i, c_i...\}
\]

- No need to resort to approximation or satisfy Markovian constraint.

\(^2\)Note: efficient caching can be performed if the Hawkes is Markov.
Simulation Statistics

We compare the simulated statistics against theoretical expectations (over 1 million simulation paths):

\[
\begin{array}{cccc}
\text{Process } m = 1 & \text{Process } m = 2 \\
\hline
\text{Time} & \text{Sim.} & \text{Expt.} & \%\text{Diff.} & \text{Sim.} & \text{Expt.} & \%\text{Diff.} \\
5.0 & 9.507 & 9.499 & 0.088 & 6.850 & 6.838 & 0.169 \\
6.0 & 9.499 & 9.499 & 0.003 & 6.844 & 6.838 & 0.078 \\
7.0 & 9.494 & 9.499 & -0.052 & 6.834 & 6.838 & -0.055 \\
8.0 & 9.507 & 9.499 & 0.087 & 6.840 & 6.838 & 0.020 \\
9.0 & 9.501 & 9.499 & 0.025 & 6.837 & 6.838 & -0.017 \\
10.0 & 9.497 & 9.499 & -0.017 & 6.837 & 6.838 & -0.019 \\
\end{array}
\]

**Figure:** Plot of simulated mean intensities vs the theoretical stationary average intensities of the three-dimensional Hawkes processes.

Verifies that our algorithm and implementation is correct.

See paper for other results.
Outline

Introduction on Hawkes Processes

Simulation of Hawkes Processes

Bayesian Inference for Hawkes
Bayesian Inference in One Slide

- Fully Gibbs sampling achieved by
  - **Auxiliary variables augmentation** – introduce additional parameters called branching structures – allow decoupling of existing parameters.
  - **Adaptive rejection sampling (ARS)** – for variables that do not have known posterior distributions, we show conditions for which the posteriors are log-concave, and sample via ARS.

- On simulated data, we demonstrate that the parameters learned using Bayesian inference is accurate and superior to MLE:

<table>
<thead>
<tr>
<th>Name</th>
<th>Var.</th>
<th>Process $m = 1$</th>
<th></th>
<th>Process $m = 2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>True</strong></td>
<td><strong>MLE</strong></td>
<td><strong>MCMC</strong></td>
<td><strong>True</strong></td>
</tr>
<tr>
<td><strong>Background intensity</strong></td>
<td>$\mu_m$</td>
<td>2.0000</td>
<td>2.0078</td>
<td>1.9026</td>
<td>1.0000</td>
</tr>
<tr>
<td>Decay rates</td>
<td>$\delta^1_m$</td>
<td>6.0000</td>
<td>6.5367</td>
<td>6.0978</td>
<td>3.0000</td>
</tr>
<tr>
<td></td>
<td>$\delta^2_m$</td>
<td>2.0000</td>
<td>2.6464</td>
<td>2.4649</td>
<td>5.0000</td>
</tr>
<tr>
<td>Shape parameters</td>
<td>$\alpha^1_m$</td>
<td>4.0000</td>
<td>4.0171</td>
<td>4.0293</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2_m$</td>
<td>2.0000</td>
<td>2.0135</td>
<td>2.0100</td>
<td>6.0000</td>
</tr>
<tr>
<td>Rate parameters</td>
<td>$\beta^1_m$</td>
<td>2.0000</td>
<td>1.9996</td>
<td>2.0193</td>
<td>4.0000</td>
</tr>
<tr>
<td></td>
<td>$\beta^2_m$</td>
<td>5.0000</td>
<td>4.9969</td>
<td>5.0426</td>
<td>3.0000</td>
</tr>
<tr>
<td>Mean square error</td>
<td>MSE</td>
<td>0.0000</td>
<td>0.1009</td>
<td><strong>0.0340</strong></td>
<td>0.0000</td>
</tr>
</tbody>
</table>

- See paper for application on modelling Dark Networks.
Summary

- Theoretical result on expected stationary intensities
- Simulation of multivariate Hawkes with superposition theory and first order statistics
- Bayesian inference on Hawkes with auxiliary variable augmentation and adaptive rejection sampling