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

[1] Presented in Monash University.

Outline

Introduction on Hawkes Processes

Simulation of Hawkes Processes

Assessments on Simulations

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.
  - For example, the number of vehicles passing an intersection in 30 minutes.

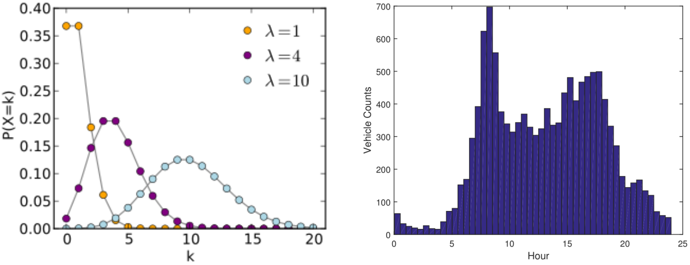


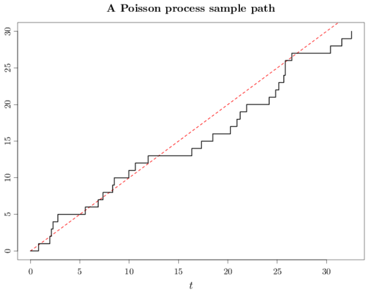
FIGURE: (left) Probability mass functions (right) Observed histogram

- Additive Property: If $X \sim \mathrm{Poi}(\lambda_1)$, $Y \sim \mathrm{Poi}(\lambda_2)$, then

$$X + Y \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$$

Background

- (Homogeneous) Poisson point 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 vehicles passing an intersection is an evolution of counts with time.



  - The number of events between two time points follow a Poisson distribution.
  - Call the evolution of counts as the counting process $N(t)$ and the times of an event happening the event times t_i.

Background

- 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. This parameter is sometimes known as intensity measure.
  - The counting process at time t follows $\mathrm{Poi}(\lambda t)$. (number of events observed until time t)
  - The difference (also called increment) in counting processes

$$N(t) - N(s) \sim \mathrm{Poi}(\lambda(t - s)) \qquad\qquad t > s$$

  - The increments for non-overlapping time points are independent.

$$N(t_4) - N(t_3) \perp N(t_2) - N(t_1) \qquad\qquad t_4 > t_3 > t_2 > t_1$$

  - The increments are time homogeneous, they depend on time difference $(t - s)$ rather than the time (t) itself.
  - Superposition property: If $N(t) \sim \mathrm{PP}(\lambda_1)$ and $M(t) \sim \mathrm{PP}(\lambda_2)$, then

$$N(t) + M(t) \sim \mathrm{PP}(\lambda_1 + \lambda_2)$$

Background

- Extension: Inhomogeneous Poisson process
  - The homogeneous Poisson process assume constant intensity λ which might not be realistic, e.g., we expect lower intensity during midnight. (see blue histogram below)
  - Instead of constant intensity, allow the intensity to vary with time: $\lambda(t)$ is a function of time.
  - Examples:
    - Piecewise linear;
    - Piecewise polynomial;
    - Cyclical functions such as sine curve.

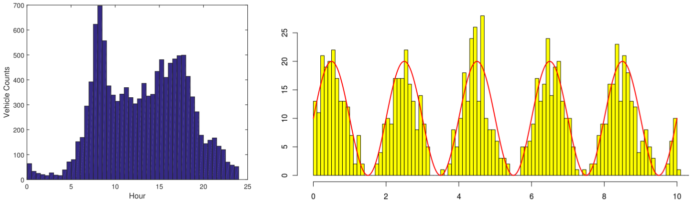


FIGURE: (left) Observed histogram (right) Generated data for IPP

Background

- Properties of inhomogeneous Poisson process
  - The counting process starts at zero: $N(t=0)=0$.
  - Parameterised intensity measure $\lambda(t)$ which is a function of time.
  - The counting process at time t follows $\mathrm{Poi}(\int_0^t \lambda(u)\,du)$.
  - The difference (also called increment) in counting processes

$$N(t) - N(s) \sim \mathrm{Poi}\left(\int_s^t \lambda(u)\,du \right) \qquad t > s$$

  - The increments for non-overlapping time points are independent.

$$N(t_4) - N(t_3) \perp N(t_2) - N(t_1) \qquad t_4 > t_3 > t_2 > t_1$$

  - The increments are no longer time homogeneous, they depend on the time (t).
  - Superposition property still holds: If $N(t) \sim \mathrm{IPP}(\lambda_1(t))$ and $M(t) \sim \mathrm{IPP}(\lambda_2(t))$, then

$$N(t) + M(t) \sim \mathrm{IPP}(\lambda_1(t) + \lambda_2(t))$$

Background

- Further extension: Intensity modulated by random processes
  - The intensity measure $\lambda(t)$ for counting process $N(t)$ can be a random function, i.e., it is a function of random processes.
  - Example:
    - Shot noise Cox process: intensity modulated by another counting process.

$$\lambda(t) = \sum_{i=1:\, t>s_i}^{M(T)} \alpha e^{-\delta(t-s_i)}$$

    - Log Gaussian Cox process: intensity modulated by exponent of a Gaussian process.

$$\lambda(t) = \exp(X(t)), \qquad X(t) \sim \mathrm{GP}$$

    - Hawkes process: intensity modulated by its own counting process – this gives self-exciting property! Simple univariate Hawkes process with exponential kernel:

$$\lambda(t) = \mu(t) + \sum_{i=1:\, t>t_i}^{N(T)} \alpha e^{-\delta(t-t_i)}$$

  - We will focus on Hawkes processes.

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):

$$\lambda(t) = \mu(t) + \sum_{i=1:\, t>t_i}^{N(T)} \alpha e^{-\delta(t-t_i)}$$

  - Decaying background intensity to capture 'edge effect'.

$$\mu(t) = \mu + Y(0)\, e^{-\delta \times t}$$

  - Random excitations:

$$\alpha \to Y_i$$
$$Y_i \sim \text{i.i.d. Gamma}$$

- 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.

Illustration of Hawkes Intensity

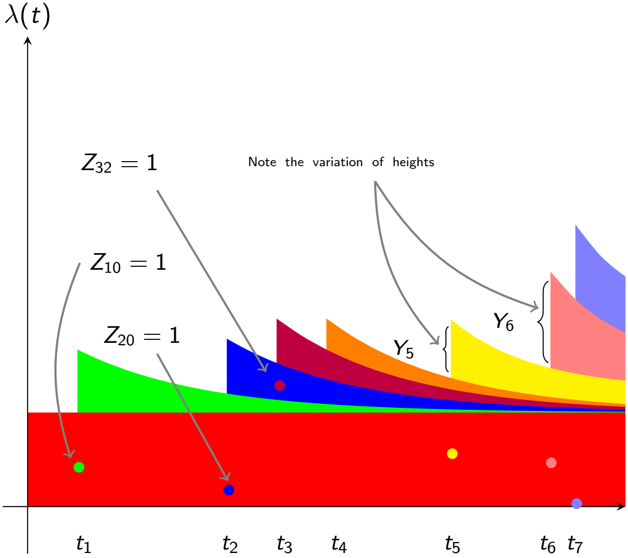


FIGURE: A sample path of the intensity function $\lambda(\cdot)$.

Multivariate Hawkes

- captures multiple event types for which the events mutually excite one another.
- Our formulation (Bivariate Hawkes):

$$\lambda_1(t) = \mu_1 + Y_1^1(0)\, e^{-\delta_1^1 t} + Y_1^2(0)\, e^{-\delta_1^2 t} + \sum_{j=1:\, t \geq t_j^1}^{N^1(t)} Y_{1,j}^1\, e^{-\delta_1^1 t} + \sum_{j=1:\, t \geq t_j^2}^{N^2(t)} Y_{1,j}^2\, e^{-\delta_1^2 t}$$

$$\lambda_2(t) = \mu_2 + Y_2^1(0)\, e^{-\delta_2^1 t} + Y_2^2(0)\, e^{-\delta_2^2 t} + \sum_{j=1:\, t \geq t_j^1}^{N^1(t)} Y_{2,j}^1\, e^{-\delta_2^1 t} + \sum_{j=1:\, t \geq t_j^2}^{N^2(t)} Y_{2,j}^2\, e^{-\delta_2^2 t}$$

  where $\lambda_1(t)$ and $\lambda_2(t)$ are the intensity functions for process 1 and 2, respectively.
- Note that the decay parameters δ are different for each process.

Illustration of Multivariate Hawkes

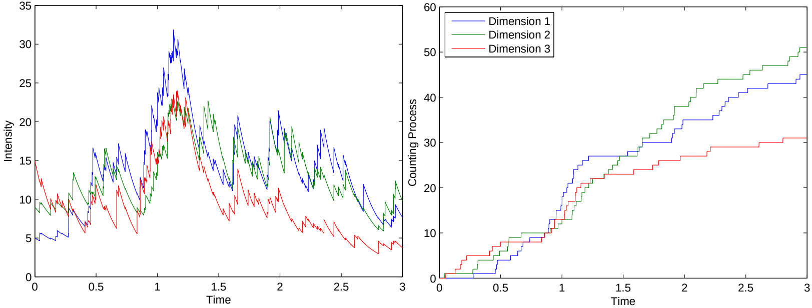


FIGURE: A sample of a three-dimensional Hawkes processes. The left plot graphs the realised intensity function $\lambda_m(t)$ for the Hawkes processes, while the right plot shows the corresponding counting processes $N^m(t)$. Each increase in the counting processes corresponds to a jump in each of the intensity functions.

Recap

- $N^i(t)$ is the number of arrivals or events of the process by time t, superscript i denote which process.
- The intensity function $\lambda(t)$ is related to the expected number of events:

$$\mathbb{E}[N(t)] = \int_0^t \lambda(u)\, du$$

- Poisson process: $\lambda(t) = const.$, arrivals of events are independent with each other, and follow the same constant rate.
- Inhomogeneous Poisson process: $\lambda(t)$ is a deterministic function of time, arrivals of events depend on the intensity function $\lambda(t)$.
- Hawkes process: $\lambda(t)$ is a function of its own counting process $N(t)$ (becomes a random function). An occurrence of an event causes 'jump' in the intensity, thereby excites more future events.
- Any question so far?

Detour: Stationarity of Hawkes process

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

$$\mathbb{E}[Y_i] < \delta$$

- The intensity process $\lambda(t)$ explodes if this condition is not satisfied:
  - When $\delta > \mathbb{E}[Y_i]$, the added intensity from an event fails to decay fast enough.
  - Causing chain reactions: intensity increases $\rightarrow$ more future events $\rightarrow$ further increases in intensity...
- Slightly more complicated condition for multivariate Hawkes (see paper for details).
- We present a theoretical result on the expected stationary intensities for our Hawkes formulation.

Outline

How to simulate a Poisson process?

- First method: simulate the counting process
  - Recall that

$$N(t) - N(s) \sim \mathrm{Poi}(\lambda(t-s)) \qquad\qquad t > s$$

  - So we can simulate an evolution of $N(t)$ sequentially by choosing $t = 0.01, 0.02, 0.03$ and so on.
  - Called grid-based method.
  - But this is not exact (an event may arrives at time $t = 0.02345$).
- Second method: simulate the event times
  - It is possible to show that the inter-arrival time Δ_i between two events follow an exponential distribution.

$$\Delta_i = t_i - t_{i-1} \sim \mathrm{Exp}(\lambda)$$

  - Once knowing Δ_i we can reconstruct t_i and $N(t)$:

$$N(t) = \sum_i 1_{t_i < t}$$

(number of events seen before time t)

How to simulate a Hawkes process?

- First method from Poisson process is difficult.
  - No nice formulation

$$N(t) - N(s) \text{ is not Poisson} \qquad t > s$$

- Second method: simulate the event times
  - Possible to derive the cumulative distribution function (cdf) of inter-arrival times:

$$\begin{aligned}\log\{1 - F(u|t_1, \cdots, t_k, \theta)\} &= -\int_{t_k}^{u} \Lambda(t|t_1, \cdots, t_k, \theta)dt \\ &= -\int_{t_k}^{u} \left\{\mu + \sum_{i=1}^{k} g(t - t_i|\theta)\right\} dt \; .\end{aligned}$$

  - Issue: cdf cannot be inverted for direct sampling.
    - Ozaki (1979) used numerical method to sample – no longer exact.
    - Dassios and Zhao (2013) recast the Hawkes process using ODE and sample the event times exactly – $\lambda(t)$ needs to be Markovian.
    - Our method: Extend DZ for Hawkes with dissimilar decays which is not Markovian – we use superposition property and first order statistics.
- Other methods:
  - Ogata's thinning method.
  - Cluster based method.

Going high level – Superposition property

- Bivariate Hawkes as example, recall intensity functions:

$$\lambda_1(t) = \mu_1 + Y_1^1(0)\, e^{-\delta_1^1 t} + Y_1^2(0)\, e^{-\delta_1^2 t} + \sum_{j=1:\, t \geq t_j^1}^{N^1(t)} Y_{1,j}^1\, e^{-\delta_1^1 t} + \sum_{j=1:\, t \geq t_j^2}^{N^2(t)} Y_{1,j}^2\, e^{-\delta_1^2 t}$$

$$\lambda_2(t) = \mu_2 + Y_2^1(0)\, e^{-\delta_2^1 t} + Y_2^2(0)\, e^{-\delta_2^2 t} + \sum_{j=1:\, t \geq t_j^1}^{N^1(t)} Y_{2,j}^1\, e^{-\delta_2^1 t} + \sum_{j=1:\, t \geq t_j^2}^{N^2(t)} Y_{2,j}^2\, e^{-\delta_2^2 t}$$

- We can treat each process as superposition of simpler processes.
- In this case, process 1 is a superposition of:
  - a homogeneous Poisson process with intensity μ_1;
  - two inhomogeneous Poisson processes with decaying intensities;
  - a self-excitation part; and
  - a shot noise Cox part.
- Sampling the inter-arrival times for each simpler process is easy and exact (cdf can be inverted).

Going high level – First order statistics

- Denote $F_A(x)$ as the cdf for a random variable A.
- If we have

$$\big(1 - F_A(x)\big) = \big(1 - F_B(x)\big)\big(1 - F_C(x)\big)$$

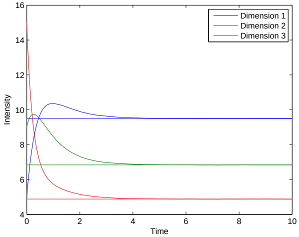
  then, $A = \min\{B, C\}$ is a first order statistics of B and C.
- Same principle holds for more than two random variables.
- We can show that the inter-arrival time for a Hawkes process is a first order statistics of the inter-arrival time of those simpler processes (details in paper).
- What this means?
  - We can sample the inter-arrival time of the simpler processes, and then take their minimum as the inter-arrival time for our Hawkes.
  - Do not need to resort to approximation or satisfy Markovian constraint.
- Additional caching techniques make our sampler efficient.

Outline

- Introduction on Hawkes Processes
- Simulation of Hawkes Processes
- Assessments on Simulations
- Bayesian Inference for Hawkes

Simulation Statistics

- We compare the simulated statistics against theoretical expectations:



	Process $m=1$			Process $m=2$		
Time	Sim.	Expt.	%Diff.	Sim.	Expt.	%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

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

- This verifies that our algorithm and implementation is correct.

Speed comparison

- We compare our simulation algorithm with several existing ones:

Sampler	Events	StDev	Time/Sim [s]	StDev [s]	Time/Event [μs]
Multivariate ($M = 50$)					
Zaatour (2014)	12 503	566	4.24	0.29	339
Our sampler	12 492	530	2.47	0.14	**198**
Univariate ($M = 1$)					
Ozaki (1979)	249 995	2505	32.73	0.69	131
Brix & Kendall (2002)	249 926	2544	14.31	1.09	57
Our sampler	249 891	2533	3.12	0.26	**13**

*Our sampler is equivalent to Dassios and Zhao (2013) for univariate Hawkes, thus not compared.

Outline

Bayesian Inference

- Fully Gibbs sampling achieved by
  - Auxiliary variables augmentation – we introduce additional parameters called branching structures that 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, which facilitates efficient sampling via ARS.
- On simulated data, we demonstrate that the parameters learned using Bayesian inference is accurate and superior to MLE:

		Process $m=1$			Process $m=2$		
Name	Var.	True	MLE	MCMC	True	MLE	MCMC
Background intensity	μ_m	2.0000	2.0078	1.9026	1.0000	1.0051	0.8555
Decay rates	δ_m^1	6.0000	6.5367	6.0978	3.0000	4.0671	3.0790
	δ_m^2	2.0000	2.6464	2.4649	5.0000	5.4443	5.2633
Shape parameters	α_m^1	4.0000	4.0171	4.0293	1.0000	1.0103	1.0076
	α_m^2	2.0000	2.0135	2.0100	6.0000	6.0907	6.0638
Rate parameters	β_m^1	2.0000	1.9996	2.0193	4.0000	4.0262	4.0407
	β_m^2	5.0000	4.9969	5.0426	3.0000	3.0223	3.0351
Mean square error	MSE	0.0000	0.1009	**0.0340**	0.0000	0.1922	**0.0148**

Real world application – Modelling Dark Networks

- What are Dark-nets?
  - Online 'anonymous' market places hidden from public access.
  - Enables buying and selling 'anything', e.g., drugs, stolen credit cards.
  - Discussion forum: Open, un-moderated, e.g., "How to avoid law".
- Why do we care?
  - Facilitate over $20b of $330b (pa) narcotic trade and nastier stuff.
  - Provide actionable intelligence & historical info.
  - Manual law enforcement processes: reactive and do not scale.

FIGURE: BlackBank site: fraudulent credit card for sale

Real world application – Modelling Dark Networks

- We model the forum posts on drugs as Hawkes process:
  - Keywords: Meth and Cannabis
  - Arrival of an event is the forum posting with relevant keywords.
  - We assume the events are mutually-exciting – people replying, update existing posts, etc.
- Results:
  - Learned parameters:

$$\begin{aligned}
&& \hat{\mu}_1 &= 0.6063\,, & \hat{Y}_1^1(0) &= 0.7202\,, & \hat{Y}_1^2(0) &= 0.7329\,,\\
&& \hat{\mu}_2 &= 0.0930\,, & \hat{Y}_2^1(0) &= 0.5166\,, & \hat{Y}_2^2(0) &= 0.5149\,,\\
\hat{\alpha}_1^1 &= 0.0599\,, & \hat{\alpha}_1^2 &= 0.1487\,, & \hat{\beta}_1^1 &= 0.0760\,, & \hat{\beta}_1^2 &= 1.4939\,,\\
\hat{\alpha}_2^1 &= 0.3894\,, & \hat{\alpha}_2^2 &= 0.1147\,, & \hat{\beta}_2^1 &= 1.9399\,, & \hat{\beta}_2^2 &= 1.4197\,.
\end{aligned}$$

  - The background intensity for the cannabis' forum (μ_1) is much higher as a result of the observation of higher posts on the cannabis forum.
  - α_m^i and β_m^i describe the distribution of the levels of excitation.
  - In this case, the expected levels of excitation are given as

$$\begin{aligned}
\hat{\mathbb{E}}[Y_{1,\cdot}^1] &= 0.7878\,, & \hat{\mathbb{E}}[Y_{2,\cdot}^1] &= 0.0995\,,\\
\hat{\mathbb{E}}[Y_{1,\cdot}^2] &= 0.2007\,, & \hat{\mathbb{E}}[Y_{2,\cdot}^2] &= 0.0808\,.
\end{aligned}$$

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
- Application on modelling Dark-nets forum data