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]Proceedings of the 8th Asian Conference in Machine Learning (ACML) 2016

Talk Outline

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.

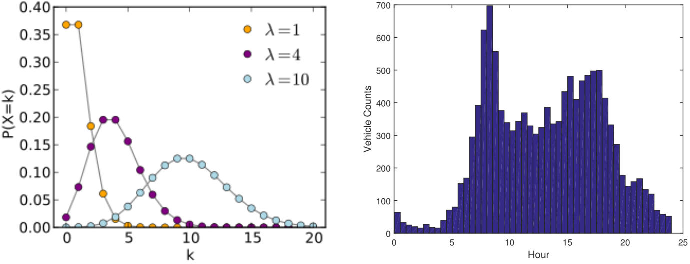


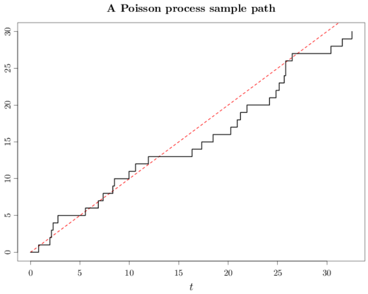
FIGURE: (left) Probability mass functions (right) Observed histogram

- 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 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:



  - 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.
  - The counting process $N(t)$ at time t follows $\mathrm{Poi}(\lambda t)$. (number of events observed until time t)
  - The difference (also called increment) in counting processes

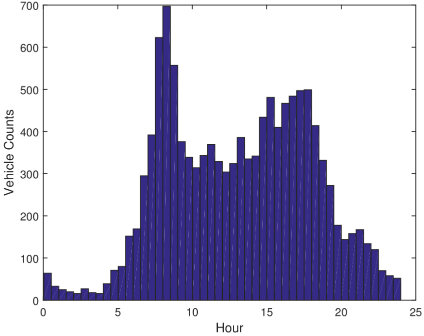
$$N(t) - N(s) \sim \mathrm{Poi}(\lambda(t - s)) \qquad t > s$$

  - Superposition property: If $N(t) \sim \mathrm{PP}(\lambda_1)$, $M(t) \sim \mathrm{PP}(\lambda_2)$, then

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

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.

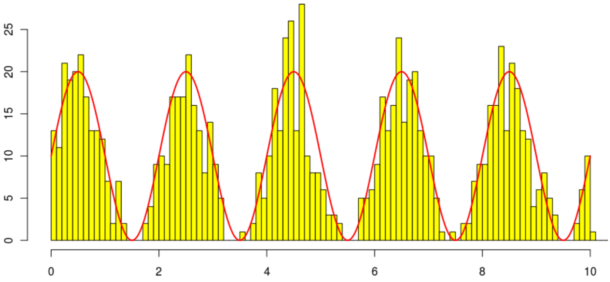


FIGURE: Generated data

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 $\mathrm{Poi}(\int_0^t \lambda(u)\,du)$.
  - The difference (also called increment) in counting processes

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

  - Superposition property still holds: If $N(t) \sim \mathrm{IPP}(\lambda_1(t))$, $M(t) \sim \mathrm{IPP}(\lambda_2(t))$, then

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

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):

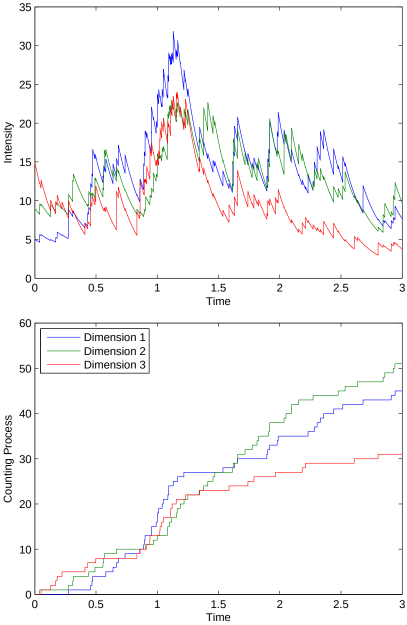
$$\lambda_1(t) = \mu_1(t) + \sum_{j=1:\, t \geq t_j^1}^{N^1(t)} Y_{1,j}^1 \, e^{-\delta_1^1 t} + \sum_{j=1:\, t \geq t_j^2}^{N^2(t)} Y_{1,j}^2 \, e^{-\delta_1^2 t}$$

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

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

- Note that the decay parameters δ 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

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

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

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

- A Hawkes with intensity $\lambda_1(t)$ is a superposition of IPP (with intensities μ_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 Δ_i for a Hawkes process is a first order statistics of these inter-arrival times:

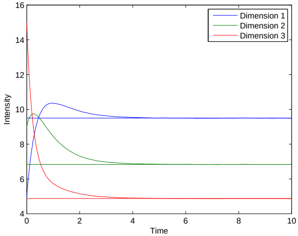
$$\Delta_i = \min\{a_i, b_i, c_i...\}$$

- 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):



	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.

- Verifies that our algorithm and implementation is correct.
- See paper for other results.

Outline

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:

		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**

- 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